# Application form

## Doctoral Grant for Teachers, round 2024-I

Please read the call for proposals and the extendable explanation carefully before you start writing your proposal.

The application form contains 'Explanatory Notes'. They can be made visible by clicking on the triangle where the arrow is pointing to. Note: on Apple computers the explanatory notes can be made visible by selecting "Outline" under the "View" tab in Word. Also, the rules and guidelines are automatically visible when you open the form in Google docs.

# 1. General information

## 1a. Title research proposal

A Neorepublican Perspective on Automated Decision-making

## 1b. Abstract of the research proposal

*Word limit: 50-100 words per summary*

### Dutch

**Een neorepublikeins perspectief op geautomatiseerde besluitvorming**

Dit beoogd proefschrift onderzoekt de legitimiteit van machine learning-profilering in institutionele besluitvorming binnen een neorepublikeins raamwerk. Het beoordeelt de politieke en morele aanvaardbaarheid, vergelijkt traditionele en op machine learning gebaseerde benaderingen, voert empirische casestudies uit en formuleert beleidsaanbevelingen. Het breidt neorepublikeins denken uit over profilering, nondominantie en machtsmisbruik, en verdiept het debat over eerlijkheid in machine learning. Vanuit theoretische en empirische analyses worden praktische tools ontwikkeld voor de verantwoorde implementatie van geautomatiseerde besluitvorming.

### English

**A Neorepublican Perspective on Automated Decision-making**

The topic of this proposed PhD research is the legitimacy of machine learning-based profiling in institutional decision-making. Grounded in a neorepublican framework, it scrutinizes when profiling raises political or moral concerns, compares traditional and machine learning-based approaches, conducts empirical case studies in different institutional domains, and proposes policy recommendations. The study contributes to and extends the neorepublican thinking on profiling in relationship to nondomination and abuse of power and deepens the debate on fairness in machine learning. After theoretical and empirical analysis, it aims to furnish practical tools to guide the implementation of automated decision-making.

# 2. Research proposal

## 2a. Introduction and research questions
*Word limit: 650 words*

Suppose someone applied for welfare at the municipality of Amsterdam somewhere between April and July 2023. In that case, chances are that their application was given a score for the risk that it was fraudulent. Based on this score, enforcement professionals could investigate the application further. A machine learning algorithm, trained on previous welfare applications, would have created the score (Gemeente Amsterdam 2023).

This type of statistical decision-making has found its way into many aspects of our lives. When you buy insurance, are trying to get credit, or are crossing a border, you are being profiled. Your particular situation is generalized to match you with people who share a set of statistically relevant characteristics. Machine learning has encouraged the adoption of these methods by adding the ability to find decision rules from large datasets (Barocas, Hardt, and Narayanan 2023, ix).

An important question with profiling, in general, is whether it isn't biased towards particular groups. Are all groups treated equally? Data-driven profiling decisions allow for measuring this bias. This has led to a flurry of activity in a young field of computer science: fairness in machine learning (Caton and Haas 2020). The field has produced essentially three mutually exclusive statistical definitions of 'fairness'. None of them is sufficient to claim fairness aligning with our moral intuitions, as they all allow for blatantly unfair practices (Barocas, Hardt, and Narayanan 2023, chaps. 3-4). Despite their shortcomings, these narrow definitions have been central in the debate about the legitimacy of machine learning-based decision-making, usually without thinking about other possible normative ideals (ibid., xi).

For example, the municipality of Amsterdam only used one specific statistical perspective on fairness when they checked their welfare fraud profiling tool for bias. By doing this, they ignored other possible fairness criteria and implicitly took fairness to just mean a lack of unlawful discrimination (Gemeente Amsterdam 2022). Their internal discussion and the resulting political debate then focused on this one particular form of bias without thinking about the legitimacy of the (automated) profiling decisions or how they might affect the power relations between the different affected parties (Groot Wassink 2023). This limited perspective is the natural consequence of the currently skewed debate: technical jargon and legal considerations are at the forefront, while ethics and political philosophy seem to lag behind.

The neorepublican theoretical framework developed by (among others) Pettit and Lovett is a political philosophy with a vast potential to fill this gap (Lovett and Pettit 2009; Pettit 1999). Its conception of political freedom as nondomination (nobody having arbitrary or unchecked control over the choices of another) serves as a fruitful ground for public philosophy. For example, neorepublicanism has made explicit recommendations on how controls around state surveillance should be institutionalized in a way that a more classic liberal idea of freedom as noninterference cannot do (Newell 2014). More recently, 'digital domination' has become a powerful lens for critically looking at big tech and making its power accountable (Hoeksema 2023; Susskind 2022).

Neorepublicanism pays special and unique attention to reigning in any potential abuse of power and arbitrary control. That focus forces us to "rethink issues of legitimacy and democracy, welfare and justice, public policy and institutional design" (Lovett and Pettit 2009, 12). Neorepublicanism can intervene forcefully on the issue of the legitimacy of machine learning-based profiling, allowing the concept of fairness in this context to be part of this rethink too. The main research question, therefore, becomes as follows:

> *How can the neorepublican framework inform normative and institutional thinking about the profiling of individuals by machine learning algorithms?*

## 2b. Potential contribution of the proposed research to science
*Word limit: 350 words*

The potential contribution to science will be threefold:

1. The neorepublican perspective on profiling will be deepened, specifically in the context of machine learning. Although this is mainly conceptual work, it will be informed by empirical research on institutionalizing decision-making in ways that contain less or no domination.

   Graf (2017) did some preliminary work looking at profiling from a neorepublican perspective. However, this work is not informed by the fairness in machine learning field, lacks the required empirical component (Lovett and Pettit 2009, 21), and is not aware of the recent work on critical or radical republicanism which adds a more structural point of view (Leipold, Nabulsi, and White 2020). In short, profiling through machine learning can profit from being scrutinized by a sharper neorepublican lens.

2. Neorepublicanism should also take on board the lessons that computer and legal science can teach us about nondomination. Through their need for formalization, algorithms force us to be explicit about what we mean when we use particular philosophical concepts. For example, the definition of privacy that is now prevalent in computer science has informed the philosophical thinking about what it means to have privacy (Kearns and Roth 2020). It is possible to do this for nondomination too: a formal (mathematically testable) definition of nondomination could be helpful in neorepublican thinking.

3. There will be a contribution from neorepublicanism to technical and legal scholarship and potentially even to actuarial science. The idea is to add neorepublicanism as a philosophical foundation to legitimize (or delegitimize) profiling decisions made with machine learning.

   The current statistical definitions of fairness are too narrowly focused on fairness as equality of opportunity and take their cue too much from the law. The discussion about fairness is undertheorized in current philosophy, with computer scientists and legal academics doing the work that political philosophers and ethicists should do (Barocas, Hardt, and Narayanan 2023, chap. 2, chap. 4, chap. 8).

# 2c. Research design, research approach and methodology
*Word limit: 1,300 words*

This will be a doctorate in philosophy, with a diverse approach in terms of research perspectives and disciplines.

In order to answer the primary research question, it is essential first to understand what a neorepublican perspective on profiling, in general, could be. Next, it is crucial to understand if there is a relevant difference between human beings doing the profiling and machine learning-based algorithms doing the same. In order to create practical tools that can help protect people from arbitrary power, it is critical to empirically understand how this type of automated decision-making is currently institutionalized. Once that is understood, it becomes feasible to chart a way forward. The research design, therefore, contains four sub-questions:

*1. When is profiling (in general) legitimate and justified from a neorepublican perspective?*

The profiling that is relevant for this research is when statistically sound generalizations that do not match with every individual are made (e.g., men are usually taller than women, but not all men are taller than all women) (Schauer 2006). When do these generalizations become worrisome from a neorepublican point of view? When and how exactly do profiling and institutional decision-making, based on these profiles, turn into domination or arbitrary control? Furthermore, under what conditions is there no risk of domination when making profiling decisions? What forms of contestation should be available to guarantee freedom in the neorepublican sense? I will catalog how the multiple 'flavors' of neorepublicanism, particularly the critical or radical turn, differ in their perspectives on this question.

This question can only be answered by looking at different practical examples of machine learning-based profiling. The examples found in this step will feed into the empirical work at the later stages of research.

*2. What is the (philosophically relevant) difference between traditional bureaucratic or institutional profiling and data-driven machine learning-based profiling?*

Not all types of automated decisions are the same. There are currently three types of automation for profiling decisions (Barocas, Hardt, and Narayanan 2023, 30):

1. Support: Taking clearly outlined decision-making rules created by humans and automating those (without machine learning). Having a machine check whether this proposal meets the formal requirements of the call for proposals would be an example of this form.
2. Replicate: Using machine learning to replicate informal profiling decisions made by humans. Trying to replicate the decisions that the committee that looks at this proposal would make (to replace the committee potentially) is an example of this form of automation.
3. Predict: Use machine learning to 'learn' profiling decision-making rules from (sufficient amounts of) data. Computers find patterns in the data that can help predict relevant outcomes, which can then inform profiling decisions. An example in the context of this application would be to train a model on all the previous proposals for this call, considering data points like whether they indeed managed to finish their PhD within five or six years and what the citation metrics of the candidate and their work are after a certain number of years. The model then predicts

which proposals have the highest chance of success according to the defined success metrics. The goal is for the computer to be better at selecting candidates than the committee would be.

From a neorepublican perspective, each of these forms of automation will come with different considerations to ensure nondomination. I will analyze all three forms, but I have a particular interest in the third form (which is becoming very prevalent, for example Amsterdam's earlier mentioned welfare application scoring algorithm).

One relevant notion in this context will be the concept of 'arbitrariness' or 'discretion' of the profiling decision, an important constituent factor of domination in neorepublican thinking (Pettit 1999; Lovett 2013). Creel and Hellman (2022) have looked at arbitrariness in the context of algorithmic decision-making, but they do not take a neorepublican perspective.

There is an interesting tension to explore here between the uniformity of automated decision-making (always treating people with the same characteristics in the same way) on the one hand and the inability of machine learning to properly deal with idiosyncratic individuals on the other hand (Alkhatib 2021). It is currently an open question which forms of automated decision-making lead to more or less domination compared to how bureaucratic institutions would make decisions before the turn to automation.

Sub-questions 1 and 2 are answered through an extensive review of relevant literature, followed by a conceptual analysis in which the methodology is to work towards a reflective equilibrium, adjusting all the principles and beliefs until they cohere with each other (Cath 2016; Rawls 1971, 48-53). The outcome of the work on these two questions will be a comprehensive neorepublican framework on profiling.

*3. What are current implicit and explicit normative considerations when bureaucracies implement systems of profiling and decision-making?*

Lovett and Pettit write that "the forms institutional protections should take must be shaped by empirical experience of the effects of those protective policies. Philosophy and theory alone cannot dictate the best way of doing things" (2009, 21). That is why this research will necessarily include empirical work.

To see how the theoretical framework developed above can translate into practical tools that can influence how these technologies are implemented, it is paramount to gain a deeper empirical understanding of the current situation and where the opportunities for a change of approach lie.

A set of three case studies, each looking at different institutional bureaucracies making machine learning-based profiling decisions in the European context, builds this understanding:

1. Risk profiling in welfare services, by a municipality.
2. Profiling at border control, by immigration or customs.
3. Predictive policing, by law enforcement.

Each of these case studies aims to understand which ideas about fairness, justice, and (non-)domination are explicit and implicit in how organizations are currently implementing decision-making through machine learning-based profiling. This understanding is built by analyzing the documentation, interviewing the relevant stakeholders, possibly observing the decision-making in action, and charting the implementation context.

By picking canonical examples of machine learning-based profiling and focusing on the transferability of insights rather than the generalizability of insights, it is possible to create knowledge that can be relevant across different application domains.

*4. What are the best ways to guarantee freedom from domination regarding machine learning-based profiling?*

The neorepublican framework developed as a result of sub-questions 1 and 2 and the empirical insights from sub-question 3 make it clear when and how machine learning-based profiling decisions are legitimate and justifiable and when it is better not to let machine learning make profiling decisions.

The intent is to make it clear what exactly are the strengths and weaknesses of the current profiling practices and then create a practical framework (for example, in the form of a decision model) that can help in deciding whether it is legitimate to use machine learning-based profiling, and if so, what controls should be institutionalized to ensure nondomination and nonarbitrariness.

The framework should be operationalized so that policymakers (drawing up regulations), lawyers, and computer scientists (or anybody else implementing and using these technologies) can use it. This practical tool will be a step towards protecting the people affected by machine learning-based profiling decisions from domination by the state and domination by the other institutions in their lives.

iHub, Radboud University's interdisciplinary research hub on digitalization and society, is an ideal environment to work on answering sub-question 4: it fosters interdisciplinary research from a wide range of disciplines. It is already deeply engaged in the urgent questions raised by society's increased digitalization and datafication.

## 2d. Potential contribution of the proposed research to society
*Word limit: 350 words*

The WRR has called AI "the new system technology". They outline how AI is leaving the lab and is being applied in society. This change can be seen in the number of patents using AI, the number of jobs in AI, the worldwide investment in AI, and the number of governments with national AI strategies. These have rapidly increased in recent years (Prins et al. 2021, 73-79).

Machine learning is one of the five types of AI currently applied in different institutional and corporate settings. The WRR report shows how machine learning as a predictive technology is used in many different domains to make decisions: in energy, in finance, in fighting fraud and criminality, at the police, in supermarkets, in media, and in advertising (ibid., 80-83).

Unfortunately, there are many examples where machine learning-based profiling has been carefully implemented but has then turned out to have unforeseen negative consequences, often discriminatory in nature (e.g., Constantaras et al. 2023). The WRR, therefore, lists five overarching tasks to embed AI in society: demystification, contextualization, engagement, regulation and positioning (ibid., 134-174).

This research will mainly contribute to what the WRR calls contextualization, "making the technology work in practice by creating an enabling socio-technical ecosystem". The focus is on AI as predictive machine learning, which differs from generative AI, like ChatGPT.

Graf (2017, 451) takes a neorepublican perspective on these issues and calls for a broader discussion of ways forward:

> *Ideally towards a practicable way to regain a collective and an individual degree of control over dynamic environments and situations that adapt to our data profiles. This discussion needs to be interdisciplinary in nature, including engineering and computer sciences as well as law and political philosophy.*

The societal goal is to use public philosophy as a program for generating positions on political issues (Lovett and Pettit 2009). The intention is to create policy proposals on regulating machine learning-based profiling and practical tools that can be used by the institutions and teams implementing these technologies.

## 2e. Contribution of the proposed research to education
*Word limit: 350 words*

Using digital technology responsibly is one of the three leading themes for the Amsterdam University of Applied Sciences (Hogeschool van Amsterdam 2022; 2023). This has created much space in education for thinking about and practicing the design of technology in its relation to society.

This research would contribute to and inform the programs in which I am currently already active: the *Business Ethics* course in the master program *Digital Driven Business*, the lectures and commissionership in the master programs *Applied AI* and *Digital Design*, and my work in the minor *Het internet is stuk, maar we gaan het repareren*. The latter is based on the Waag Futurelab's thinking around the *public stack* (Van der Waal et al. 2020). A neorepublican take on machine learning-based decision-making can enrich the foundational layer in this public stack.

There will also be opportunities for students to be involved in my research, both in the empirical phase (analyzing documents, for example) and in the final stages, where the clarity of the conceptual analysis can be supported by graphic and informational design. I already have experience supervising students in these activities through my current work as a researcher in the *Responsible IT* research group. I plan to further develop my ability to connect research to my teaching practice.

Finally, I want to generalize my findings and translate them into educational practices that can broadly appeal throughout the (vocational) education domain. Any practical designs and tools made from my research can be applied by students working on technologies through a responsible lens from many disciplines. One way of raising awareness around my work is through a publication in a journal (or magazine) for (higher) education like *Het Onderwijsblad*, *Tijdschrift voor Hoger Onderwijs*, or *Onderzoek van Onderwijs*.

## 2f. Literature references

*Maximum: 30 references*

Alkhatib, Ali. 2021. "To Live in Their Utopia: Why Algorithmic Systems Create Absurd Outcomes." In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–9. CHI '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445740.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan. 2023. Fairness and Machine Learning: Limitations and Opportunities. The MIT Press. https://mitpress.mit.edu/9780262048613/fairness-and-machine-learning/.

Cath, Yuri. 2016. "Reflective Equilibrium." In The Oxford Handbook of Philosophical Methodology, edited by H. Cappelen, T. Gendler, and J. Hawthorne, 213–30. Oxford University Press.

Caton, Simon, and Christian Haas. 2020. "Fairness in Machine Learning: A Survey." arXiv. https://doi.org/10.48550/arXiv.2010.04053.

Constantaras, Eva, Gabriel Geiger, Justin-Casimir Braun, Dhruv Mehrotra, and Htet Aung. 2023. "Inside the Suspicion Machine." Wired, March 6, 2023. https://www.wired.com/story/welfare-state-algorithms/.

Creel, Kathleen, and Deborah Hellman. 2022. "The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision-Making Systems." Canadian Journal of Philosophy 52 (1): 26–43. https://doi.org/10.1017/can.2022.3.

De Zwart, Hans. 2017. "The Future Is False Positive." Presented at the RightsCon 2017, Brussels, March 28. https://talks.hansdezwart.nl/rightscon-2017/#/.

———. 2018. "Freedom From Domination and Our Technological Predicament." Presented at the Amsterdam Privacy Conference (APC), Amsterdam. https://philpapers.org/rec/DEZFFD.

Gemeente Amsterdam. 2022. "Bias Analyse WPI Algoritme."

———. 2023. "Onderzoekswaardigheid: Slimme Check Levensonderhoud." Amsterdam Algoritmeregister. 2023. https://algoritmeregister.amsterdam.nl/ai-system/onderzoekswaardigheid-slimme-check-levensonderhoud/1086/.

Graf, Eike. 2017. "When Automated Profiling Threatens Our Freedom: A Neo-Republican Perspective." European Data Protection Law Review (EDPL) 3: 441. https://heinonline.org/HOL/Page?handle=hein.journals/edpl3&id=480&div=&collection=.

Groot Wassink, Rutger. 2023. "Brief Pilot Digitalisering Fraudepreventie." https://amsterdam.raadsinformatie.nl/document/12521370/1/09012f9780bd52db.

Hoeksema, Bernd. 2023. "Digital Domination and the Promise of Radical Republicanism." Philosophy & Technology 36 (1): 17. https://doi.org/10.1007/s13347-023-00618-7.

Hogeschool van Amsterdam, dir. 2022. Professor of Practice - Marleen Stikker over Digitalisering. https://www.youtube.com/watch?v=1Ye8S3Ke2RE.

———. 2023. "Digitalisering." May 1, 2023. https://www.hva.nl/over-de-hva/wie-wij-zijn/digitalisering/digitalisering.html.

Jacobs, Bart. 2022. "Republicanisme Als Modern Europees Verhaal." In 77, iteratuur en recht. Liber Amicorum voor prof. dr. T.J.M. Mertens:67–77. Deventer: Wolters Kluwer. https://hdl.handle.net/2066/250795.

Kearns, Michael, and Aaron Roth. 2020. The Ethical Algorithm: The Science of Socially Aware Algorithm Design. Oxford, New York: Oxford University Press.

Leipold, Bruno, Karma Nabulsi, and Stuart White, eds. 2020. Radical Republicanism: Recovering the Tradition's Popular Heritage. Oxford, New York: Oxford University Press.

Lovett, Frank. 2013. A General Theory of Domination and Justice. Oxford, New York: Oxford University Press.

Lovett, Frank, and Philip Pettit. 2009. "Neorepublicanism: A Normative and Institutional Research Program." Annual Review of Political Science 12 (1): 11–29. https://doi.org/10.1146/annurev.polisci.12.040907.120952.

Newell, Bryce Clayton. 2014. "Technopolicing, Surveillance, and Citizen Oversight: A Neorepublican Theory of Liberty and Information Control." Government Information Quarterly 31 (3): 421–31. https://doi.org/10.1016/j.giq.2014.04.001.

Pettit, Philip. 1999. Republicanism: A Theory of Freedom and Government. Oxford University Press. https://doi.org/10.1093/0198296428.001.0001.

Prins, Corien, Haroon Sheikh, Erik Schrijvers, Eline De Jong, Monique Steijns, and Mark Bovens. 2021. "Opgave AI. De nieuwe systeemtechnologie." Rapport. Ministerie van Algemene Zaken. November 11, 2021. https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie.

Rawls, John. 1971. A Theory of Justice: Original Edition. Harvard University Press. https://doi.org/10.2307/j.ctvjf9z6v.

Schauer, Frederick. 2006. Profiles, Probabilities, and Stereotypes. Cambridge, MA: Belknap Press.

Susskind, Jamie. 2022. The Digital Republic. Pegasus Books. https://www.simonandschuster.com/books/The-Digital-Republic/Jamie-Susskind/9781643139012.

Van der Waal, Sander, Marleen Stikker, Max Kortlander, Quirine Van Eeden, Tom Demeyer, and Stefano Bocconi. 2020. "European Digital Public Spaces." https://waag.org/sites/waag/files/2021-04/Waag%20Report%20on%20Digital%20European%20Public%20Spaces.pdf.

# 2g. Work plan

*Maximum: 2 pages*

| Year | Work plan 2024-2030 (based on 5 years of funding after the decision date, with 1 year extension without funding) |
|------|------|
| 2024/25 | • Present my research plan to the team at iHub in Nijmegen and relevant research groups at the Amsterdam University of Applied Sciences.<br><br>• Do an extensive literature review of fairness in machine learning, fairness in actuarial science, radical/critical republicanism, neorepublicanism in the digital domain, the moral justifications for profiling, and (political) philosophy that discusses profiling in terms of nondomination.<br><br>• Do the required courses at Radboud's Graduate for the Humanities (on scientific integrity and data management).<br><br>• Attend a workshop/conference on neorepublicanism, preferably in the context of technology—for example, the *Republics & Republicanism Biennial*.<br><br>• Update the research plan based on of the feedback, the literature review, the conference/workshop, and the courses.<br><br>• Start drafting a first article with a neorepublican perspective on profiling, based on the answers to sub-question 1. |
| 2025/26 | • Finalize the first article.<br><br>• Start drafting a second article with a neorepublican perspective on automated decision-making, based on the answers to sub-question 2.<br><br>• Visit the 2025 ACM Conference on *Fairness, Accountability, and Transparency (FACT)*, a conference about fairness in machine learning.<br><br>• Organize access to the documentation and stakeholders for the three case studies.<br><br>• Finalize the second article.<br><br>• Enroll in the *Doing Case Study Research* course at the Radboud Summer School. |
| 2026/27 | • Present the first and/or second article at a relevant conference.<br><br>• Do the three case studies, noting potential lessons for a practical tool.<br><br>• Start drafting a third article on the implicit and explicit thinking about fairness, justice, and (non-)domination inside different institutional bureaucracies and draw relevant lessons for the neorepublican research program.<br><br>• Organize a workshop with stakeholders involved in institutionalized machine learning-based profiling to kick-start the thinking on practical tools.<br><br>• Start drafting a fourth article about practical tools from a neorepublican perspective that can help assess and improve the legitimacy of automated decision-making. |
| 2027/28 | • Finalize the third article.<br><br>• Finalize the fourth article. |

| | |
|---|---|
| | • Present the third and/or fourth article at a conference.<br>• Plan the finalization of my dissertation. |
| 2028/29 | • Integrate the four articles into a final dissertation.<br>• Defend my dissertation. |

## 2h. Data management section

1. Will data be collected or generated that are suitable for reuse?

☑ Yes: Please answer questions 2 to 4 below.

☐ No: Then explain why the research will not result in reusable data or in data that cannot be stored or data that for other reasons are not relevant for reuse **Please note:** NWO will not necessarily accept that data will not be suitable for reuse.

Type here

2. Where will the data be stored during the research?

I will use the Amsterdam University of Applied Sciences research data management infrastructure. This means that the data will only be stored in the following three places:

• Locally on my laptop (encrypted at rest).
• At the institution's Microsoft OneDrive installation.
• At the institution's figshare installation.

3. After the project has been completed, how will the data be stored for the long-term and made available for the use by third parties? For whom will the data be accessible?

Any empirically collected data will be stored in the Amsterdam University of Applied Sciences' figshare installation for the long term. Any data that allows for it (from a data protection perspective) will be available openly under a Creative Commons license. Data that should be kept confidential can only be accessed by me and the figshare administrators, including the Data Steward of the Faculty of Digital Media and the Creative Industries.

4. Which facilities (ICT, (secure) archive, refrigerators or legal expertise) do you expect will be needed for the storage of data during the research and after the research? Are these available?

No other facilities will be necessary outside of those that the Amsterdam University of Applied Sciences can readily provide within their standard research data management infrastructure. If this were to change, I have access to a dedicated Data Steward at of the Faculty of Digital Media and the Creative Industries Knowledge Centre. They will be able to connect me to all the necessary knowledge and tools.

## 2i. Ethical aspects

| | Not applicable | Not yet applied for* | Applied for* | received* |
|---|---|---|---|---|
| Approval from the Medisch Ethische Toetsingscommissie (METC) | ☑ | ☐ | ☐ | ☐ |
| Approval from an animal experiments committee (DEC) | ☑ | ☐ | ☐ | ☐ |
| Permission for research granted under the population screening Act (WBO) | ☑ | ☐ | ☐ | ☐ |

*When applicable, you must send a copy of the approval/permission after your proposal has been granted before you can start your project.*

# 3. Budget

## 3a. Duration of the project

Planned starting date (start of funding)

01-09-2024

## 3b. Category of educational establishment in which you are employed

☐ Primary education  ☐ Further professional education

☐ Special education  ☑ Higher professional education

☐ Secondary education

## 3c. Current contract (in fte)

| 1. | Permanent position (indefinite) at the moment of the deadline | fte (min 0.4) |
|---|---|---|
| | Hogeschool van Amsterdam | 0,8 |
| 2. | Other positions (permanent/temporary) at the moment of the deadline | fte |
| | N/a | n/a |
| 3. | Total | |
| | Total fte's in contract(s) | 0,8 |

## 3d. Description current position

*Maximum: 1 page, including the table*

My home base at the Amsterdam University of Applied Sciences is the *Communication and Multimedia Design* (CMD) program, where I teach most. I teach in the faculty-wide minor *Het internet is stuk, maar we gaan het repareren* (The internet is broken, but we are going to fix it), and I supervise students who are graduating with their final projects. In the Digital Driven Business program, I am solely responsible for teaching a course titled *Business Ethics*.

Part of my time, I am seconded to the *Responsible IT* research group, providing a philosophical (often ethical) angle to the group's research. The most recent research project that I finished is *Als de machine kiest: Het gebruik van kunstmatige intelligentie in werving en selectie* (When the machine decides: the use of artificial intelligence in recruitment). Currently, I am drafting a paper on the lessons I have learned from being inside the team running Amsterdam's Algorithm Registry.

About ten percent of my time in the research group is spent teaching about the research at different programs throughout the institute, for example, in the *Platform Stone* of *Creative Business*, in the Master *Digital Design*, and in the Master *Applied AI*.

| 1. | Teaching tasks | Share in % |
|----|----------------|------------|
|    | Teaching (in class) | 45% |
|    | Other teaching-related tasks: preparing class, meetings related to teaching, checking, supervision | 10% |
| 2. | Non-teaching-related tasks (excluding research) | |
|    | N/a | n/a |
| 3. | Research task (when applicable) | |
|    | Researcher at the Responsible IT research group | 45% |
| 4. | Total | 100% |

## 3e. Extent of exemption requested (in fte)

0,4 fte

# 4. Curriculum Vitae

## 4a.  Motivation for doctorate research
*Word limit: 350 words*

Throughout my career, I have focused on how technology impacts society (for the good and the bad), as evidenced by my work as an educational technologist in the corporate and higher education sectors and as a director of Bits of Freedom, where I fought for digital rights. Currently, I am a lecturer/researcher at the HvA and a co-founder and volunteer of the Racism and Technology Center, where I continue my work in promoting social justice.

I stumbled into neorepublicanism as a philosophy when I quit using Google's products and services and was looking for an explanation as to why I literally felt liberated, even though I had only managed to limit my options. A classic liberal outlook on freedom (freedom as noninterference) had way less explanatory power for this fact than a neorepublican perspective (freedom as nondomination): by getting rid of Google, I had also gotten rid of its arbitrary power to dominate me (De Zwart 2018). Bart Jacobs came into view as a promotor when I found out that he had come to very similar conclusions about the potential role of neorepublicanism in our digital world (Jacobs 2022).

Over the past years, one of my research interests has been machine-based (or automated) decision-making, particularly probabilistic profiling through machine learning. While this technology has become ubiquitous in society and has some genuine benefits, it is not without flaws. In my 2017 talk at RightsCon, titled *The Future is False Positive* (De Zwart 2017), I explained how we can all expect to be classified wrongly at some point. The fact that, unfortunately, some of us can expect this to happen more often than others is a strong motivation for me to work on making progress in the thinking about profiling.

With a solid understanding of technology and a background in political philosophy (a relatively rare combination), I feel well equipped – not to say duty-bound – to contribute to this debate. I get a lot of pleasure and energy from the sustained intellectual rigor that this type of philosophical research requires, and I am very much looking forward to it.

## 4b.  Relevant activities and output
*Word limit: 300 words*

Relevant previous and current roles:

- Researcher (and ethicist) at the Responsible IT research group at the AUAS (2020 - now)

- Co-founder and board member of the Racism and Technology Center (2021 - now)

- Executive Director at digital rights organization Bits of Freedom (2013-2019)

Relevant research:

- Als de machine kiest (with Pascal Wiggers) – Research report about using AI in recruitment processes.

- [Kort advies over de 'Slimme check levensonderhoud'](#) – Negative advice for the Amsterdam City Council about the use of a risk profiling algorithm in the context of welfare.

- [Freedom From Domination and Our Technological Predicament](#) – Conference paper presented at the 2018 Amsterdam Privacy Conference.

- [Freedom and Justice in our Technological Predicament](#) – Master thesis on neorepublicanism and big tech.

Relevant other activities and other writings:

- Lead learning designer of the minor titled *The internet is broken, but we are going to fix it* – Teaching students to think critically about technology in relation to public values like equal treatment and justice.

- Editor and co-author of the 4-weekly newsletter of the Racism and Technology Center, including texts on the potential discriminatory effects of automated profiling. E.g.:

    - [Current state of research: Face detection still has problems with darker faces](#)

    - [Algorithm to help find fraudulent students turns out to be racist](#)

    - [Representing skin tone, or Google's hubris versus the simplicity of Crayola](#)

    - [Racist Technology in Action: Rotterdam's welfare fraud prediction algorithm was biased](#)

- Many speeches and interviews on human rights and technology (often as a director of Bits of Freedom), for example:

    - [Ai Weiwei is Living in Our Future](#)

    - ['In de Tweede Wereldoorlog hadden we wél wat te verbergen'](#)

    - My podcast feed (listen.hansdezwart.nl) with appearances on podcasts and radio, and my personal list of writings and interviews (insights.hansdezwart.nl/author/hans-de-zwart).

# 5. Recommendation from the promotor and co-promotor/daily supervisor(s)

## 5a. Details of intended supervisor (promotor)

| | |
|---|---|
| Title(s): | Prof.Dr. |
| Initial(s): | B.P.F. |
| Prefix: | |
| Last name: | Jacobs |
| Expertise (in keywords): | Security, Privacy, Identity, Logic, Probability |
| Holding ius promovendi since *(only to be provided in case the promotor is an associate professor (UHD)* | |
| University/Institute: | iHub (interdisciplinary research hub on digitalization and society), Radboud University Nijmegen |
| Address for correspondence: | Erasmusplein 1 |
| ZIP code: | 6525 HT |
| City: | Nijmegen |
| Telephone number: | +31 24 3652236 |
| E-mail address: | bart@cs.ru.nl |

## 5b. Daily supervision

| | |
|---|---|
| Who is/are in charge of the daily supervison? | ☑ Responsible supervisor **(**Promotor) <br> ☐ Supervisor (Co-promotor) <br> ☐ Daily supervisor |

## 5c. Details of the co-promotor(es)/daily supervisor(s)

| | |
|---|---|
| Title(s): | Dr. |
| Initial(s): | G. |
| Prefix: | |
| Last name: | Treiber |
| Expertise (in keywords): | Political Philosophy, including Republicanism |
| University/Institute/Educational establishment: | iHub (interdisciplinary research hub on digitalization and society), Radboud University Nijmegen |
| Role | ☑ Supervisor (Co-promotor) <br> ☐ Daily supervisor |

| Title(s): | Dr. |
|---|---|
| Initial(s): | M.J. |
| Prefix: | |
| Last name: | Becker |
| Expertise (in keywords): | Ethics, including Ethics of Digitalization |
| University/Institute/Educational establishment: | iHub (interdisciplinary research hub on digitalization and society), Radboud University Nijmegen |
| Role | ☑ Supervisor (Co-promotor)<br>☐ Daily supervisor |

# 5d. Recommendation of the responsible supervisor (promotor)/daily supervisor(s)

Please read the expandable explanatory notes for this section carefully and pass them on to your intended promotor.

The questions below can be answered by the promotor, the co-promotor/daily supervisor, or by all in close collaboration.

### 1. Quality of the candidate for PhD-research

*(Please consider: prior education of the candidate in relation to the proposed research, academic skills, motivation of the applicant.)*
Hans de Zwart has a master's in philosophy and is thus well-prepared for the proposed normative research. He has been very active in this area for many years, so he knows the field well and thus has a sound basis for the proposed empirical research. Over the past years, Hans has regularly published (popular) philosophical articles on digitalization. He is a sharp thinker, and he writes well. He is in an ideal position to carry out the proposed research.

### 2. Collaboration with the candidate

*(Please consider: how did the collaboration come into existence, how were tasks divided between the applicant and the supervisor(s) when writing the research proposal?)*
I have studied mathematics and philosophy, and my academic career focused on computer science. This work became broader and more interdisciplinary during the last decade, including legal and ethical/political philosophy. In this context, I also came across republicanism as a fruitful normative view on freedom that better suits digitalization than liberalism.

### 3. Potential contribution of the dissertation to the line of research of the surpervisor(s) and to science in general

I already supervise one PhD student on republicanism, Bernd Hoeksema, together with my colleague Ronald Tinneveld, a professor in the philosophy of law. I am happy to supervise Hans de Zwart as another PhD student in this area, with his own angle (machine learning decisions). It will strengthen and broaden the (local) research on republicanism.

### 4. Feasibilty of the proposed research (max 6 years in total, of which the 5 first years will be funded)

*(Please consider: effectiveness of the approach, scale of the research, support with conducting the research by others, filling methodological gaps in the applicant's knowledge)*
Since Hans de Zwart comes well-prepared and arrives in a multidisciplinary research center (iHub) with many people working on relevant topics, the proposed research can quickly take shape. Finishing a PhD thesis in 5 years is feasible, also because Hans has ample writing (and presentation) experience. Successful termination is highly likely since the proposed research has a clear focus and trajectory.

### 5. Composition of the supervising team/information on supervision

*(Please consider: who will supervise the applicant? Will the applicant be embedded in an existing research group/Graduate School? Will there be regular meetings between the applicant and the supervisor(s)? How does the expertise of the supervisor(s) relate to the subject matter/methodology of the proposed research?)*

The plan is that Hans will spend one day per week (physically) at iHub in Nijmegen within an active research environment. Two other supervisors will be closely involved, forming a tight supervision team with me, Bart Jacobs, as the responsible supervisor.

### 6. Education/training plan

*(Please consider: intended training (potentially in addition to the regular course load of the Graduate School), what kind of skills does the applicant need to acquire for this particular research?)*

Hans will enroll in Radboud's Graduate School for the Humanities with mandatory training courses in scientific integrity and research data management. Enrollment in the Doing Case Study Research in the Social Sciences course of the Radboud Summer School will support his empirical work.

# 6. Administrative details

## 6a. Personal details of the applicant

| | |
|---|---|
| Title(s): | MA |
| Initial(s): | H. |
| Prefix: | de |
| Surname: | Zwart |
| Preferred language for correspondence with NWO: | ☐ Dutch<br>☑ English |
| Address for correspondence (where you can be reached during the assessment procedure): | Van Zeggelenplein 21 |
| ZIP code: | 2032 KA |
| City: | Haarlem |
| Telephone: | +31 6 2185 6845 |
| E-mail address: | h.de.zwart@hva.nl |
| Name and address of the school/establishment | Hogeschool van Amsterdam<br>Wibautstraat 3b<br>1091 GH  Amsterdam |

## 6b. Master's degree at a university

| | |
|---|---|
| University: | University of Amsterdam |
| Faculty/Department: | Faculty of Humanities |
| Discipline: | Philosophy |
| Date of graduation: | 31-08-2018 |
| Major: | n/a |
| Title MA thesis: | Freedom and Justice in our Technological Predicament |
| Assessment MA thesis: | 8,0 |

## 6c. Degree higher education or scientific BA

| | |
|---|---|
| University/College for Higher Education: | University of Amsterdam |
| Faculty/Department: | Faculty of Humanities |
| Discipline: | Philosophy |
| Date of graduation: | 28-08-2017 |
| Major: | n/a |
| Titel graduation thesis: | Liberty, Technology and Democracy |
| Assessment graduation thesis: the | 9,0 |

## 6d. Other degrees (higher education or academic)

University/College for Higher Education: Amsterdam University of Applied Sciences

Faculty/Department: Faculty of Sports and Nutrition

Discipline: Academy for Physical Education

Major: Leraar voortgezet onderwijs 1e graad lichamelijke opvoeding

Date of graduation: 31 August 2000

## 6e. Other relevant information regarding your education

Not applicable

# 7. Declaration and signature

## 7a. Have you requested funding for this research elsewhere or do you already receive funding/research time for this research (from your school or from another funder)?

| ☑ No | |
|---|---|
| ☐ Yes | Please provide the requested details |

## 7b. Have you already received public funding for PhD research in general (regardless the subject)?

| ☑ No | |
|---|---|
| ☐ Yes | Please provide the requested details |

## 7c. Signature of the applicant

*Please note: you must tick all boxes.*

| ☑ | By submitting this form via ISAAC, I declare that I have completed this form truthfully and completely. |
|---|---|
| ☑ | By submitting this application I declare that I satisfy the nationally and internationally accepted standards for scientific conduct as stated in the *Netherlands Code of Conduct for Research Integrity* (2018). More information about the code of conduct and the policy of NWO on research integrity can be found on the website: www.nwo.nl/integrity. |

| Signature | |
|---|---|
| Date | 22-01-2024 |