THE AUSTRALIAN NATIONAL UNIVERSITY

Position Description

College/Division:	College of Arts and Social Sciences
Faculty/School/Centre:	Research School of Social Sciences
Department/Unit:	School of Philosophy
Position Title:	Research Fellow
Classification:	**Academic Level B**
Position No:	
Responsible to:	Professor Seth Lazar

PURPOSE STATEMENT:

The ANU is seeking world-class researchers to join a team of philosophers, computer scientists, lawyers and social scientists, on the Humanising Machine Intelligence (HMI) Grand Challenge project. Our goal is to contribute to the design and adoption of sociotechnical systems necessary for democratically legitimate AI. Knitting together insights from computer science, law, philosophy, political science, and sociology, HMI will help shape government regulation of AI systems; enable industry practitioners to develop AI systems that comply with, and exceed, those regulatory standards; and shape an international research community that supports those two goals.

KEY ACCOUNTABILITY AREAS:

Position Dimension & Relationships:

We seek to appoint a researcher who can progress our research programmes, which include work on the following themes: *Automating Governance*, *Personalisation*, *Algorithmic Ethics*, *Ethics of Human-AI Interaction*, and *Philosophy of Data Science and AI*.

Automating Governance focuses on how the state and state-like entities use data and AI to exercise power over people. Our goal is to identify the risks and opportunities associated with these practices—an exercise in legal and moral diagnosis—then to understand what we *should* be aiming at, and then to design sociotechnical systems that achieve these objectives. Within this thematic area, one stream of research focuses on the implications of AI for public law, the other focuses on the broader question of how data and AI lead us to rethink questions about the authority of states and state-like entities (such as digital intermediaries).

Our ***Personalisation*** subtheme focuses on the ways in which non-state actors use data and AI to shape our online lives around our revealed interests and behaviours, in order to hold our attention and influence our behaviour. It falls into three research streams, focused on Algorithmic Amplification, Automated Influence, and Bias and Discrimination. Our Algorithmic Amplification stream includes both computational and qualitative research aimed at understanding the ways in which recommender systems direct attention around digital platforms, and on the economics of online attention. We are also working on normative projects on how fairness constrains the distribution of online attention, as well as precisely what the algorithmic amplification of online speech should aim at. On Automated Influence, we are addressing the value of privacy, the inadequacy of existing data protection regulations, the political philosophy of recommender systems and online behavioural advertising, and the goal of developing privacy- and fairness-preserving recommender systems. On Bias and Discrimination we are exploring how fairness as understood in the machine learning literature maps onto fairness in regulatory instruments like the GDPR, as well as enriching the debate by bringing in perspectives from sociology and from theories of structural discrimination, at the same time as developing algorithmic tools for the use of data and AI in industry that achieve promised social benefits without exacerbating disparate impacts.

Our ***Algorithmic Ethics*** subtheme notes that if we're going to design AI systems that reflect our values as democratic societies, we have to figure out how to either train AI systems to learn normative goals and constraints, or else encode those goals and constraints into those systems. Either approach presupposes that it is possible to represent complex normative theories in terms that are computationally accessible and infer optimal decisions that comply with these normative theories in a computationally tractable manner. We want to determine whether this is possible. We are therefore working on translating formal representations of moral theories into computational languages, and assessing their degree of complexity, as well as the other requirements for operationalising them in realistic contexts. We are also pursuing foundational work in AI, for example on integrating symbolic and learning approaches to AI in order to realise value-aligned AI systems that are more trustworthy and easier to explain. And we are building on these insights in the design of robots and other autonomous systems, developing standards for evaluating the safety of autonomous vehicles in collaboration with the Assuring Autonomy International Programme in the UK, showing how

different moral theories can be represented by AVs, and developing new approaches to the design of strategically compassionate robots.

The Ethics of Human-AI Interaction, starts from the premise that in the design of data and AI systems we *must* take into account the predictable ways in which people will use or misuse those systems, and in particular the predictable cognitive biases which will shape our misuse. We must also attend to the ways in which using new technologies reshapes us, as people—the hammer shapes the hand. This involves empirical work considering how, for example, we systematically misattribute responsibility when we work in human-machine teams, as well as how we irrationally defer to automated systems even when we have sufficient reason to be sceptical of their veracity. We are also exploring how working in human-machine teams impacts on our capacity for moral behaviour and judgment, considering in particular how our practices of role-taking change when we work alongside AI systems, as well as the social implications of outsourcing decisions that require us to exercise moral judgment to automated systems.

The central idea of the **Philosophy of Data Science and AI** is to apply the methods of the philosophy of the sciences to the topic of data science and AI. Historically, work in this vein has been limited to philosophers of mind, who have, until recently, often operated with outdated understandings of the state of the art in AI research. This new field is undertaken by philosophers who are deeply immersed in current AI research, and are identifying therein novel contributions ranging from conceptual analysis, to first-order contributions to computational theory, to exposing analogies between central elements of data science and other theses in the philosophy of the sciences. Like work in other areas of philosophy of the sciences, work in this subfield has the potential to both illuminate data science and AI for computer scientists, and to make first-order philosophical advances.

The HMI project chief investigators are: Professors Seth Lazar (Project Leader) and Colin Klein, and Associate Professor Katie Steele (Philosophy), Professors Sylvie Thiébaux and Lexing Xie, and Associate Professor Hanna Kurniawati (Computer Science), Dr. Jenny Davis (Sociology), Professor Toni Erskine and Dr Sarah Logan (Political Science), Associate Professor Will Bateman and Dr. Damian Clifford (Law).

We are looking for a 24 month postdoctoral fellow to advance one or more of these projects, or to pursue a closely related research programme. Our primary criterion is demonstrated research excellence in a discipline area relevant to the project, and the clear potential to make internationally-recognised progress on these and related themes. An interdisciplinary background is not required, but successful applicants will be ready and equipped to engage with scholars from other disciplines. We are keen to appoint someone who can collaborate with lawyers on the HMI team, though this is not a strict requirement.

Successful applicants will publish internationally influential research in leading peer-reviewed venues (as suited to their discipline). We expect them to go on from the ANU to leading positions in academia and industry. A crucial component of their role will be to help maintain and build the HMI community at ANU and globally, through active participation in the collective research life of the project, and service roles such as convening a seminar series and international workshops.

The position is open with respect to field. A prior track record of work on AI and society (and related issues as appropriate to their training) is desirable but not required. The successful applicant will be expected to begin the role in a position to work on these themes, and to focus their work on topics that advance the HMI research agenda.

This position will be attached to the HMI project, and will be supervised by one of the project executive team (Lazar, Davis, Xie, Kurniawati, Erskine) on behalf of the project executive as a whole, and based in the corresponding school. We strongly encourage anyone who meets the selection criteria to apply, regardless of disciplinary background.

We strongly encourage applications from candidates from backgrounds that have historically been unrepresented in their field.

Role Statement:

Specific duties required of a **Level B Research Intensive Academic** may include:

- Undertake research that contributes to the goals of the HMI project, independently and as part of a team, with a view to: publishing original, innovative, and high impact research in world-leading refereed journals and conference proceedings; presenting research at academic seminars and at national and international conferences; and collaborating with other researchers at a national and international level.
- Build the HMI research community by helping organise regular seminars, reading groups, and workshops, and actively participating in community activities including virtual and in person.
- Contribute, at a restricted intensity, to discipline-appropriate teaching activities within the University at the undergraduate and graduate levels including honours supervision.
- Supervise less senior academic staff and research support staff in their research area.
- Assist in outreach activities including to prospective students, research institutes, industry, government, the media and the general public.
- Maintain high academic standards in all education, research and administration endeavours.
- Take responsibility for their own workplace health and safety and not wilfully place at risk the health and safety of another person in the workplace.
- Other duties as required consistent with the classification level of the position.

Skill Base
A **Level B Academic** will normally have completed a relevant doctoral qualification or have equivalent qualifications or research experience.

In addition they may be expected to have had post-doctoral research experience that has resulted in publications, conference papers, reports or professional or technical contributions that give evidence of research ability.

SELECTION CRITERIA:

1. PhD in one of the following at the time of appointment: philosophy, computer science, law, mathematics, economics, political science, sociology, engineering, or another relevant discipline.
2. Demonstrated capacity to pursue research at the highest levels of international scholarship in their discipline.
3. Evidence of the ability to articulate and prosecute innovative research in the field of the HMI project and a vision for the activities they will undertake at the ANU.
4. Capacity to engage in cross-disciplinary research collaboration and build a research community.
5. Ability and willingness to teach (at a restricted intensity) at all levels.
6. The ability to supervise and graduate high quality PhD/Masters/Honours research students
7. Excellent oral and written English language skills and a demonstrated ability to communicate and interact effectively with a variety of staff and students in a cross-disciplinary academic environment and to foster respectful and productive working relationships with staff, students and colleagues at all levels.
8. A demonstrated high-level understanding of equal opportunity principles and a commitment to the application of these policies in a University context.

The ANU conducts background checks on potential employees, and employment in this position is conditional on satisfactory results in accordance with the Background Checking Procedure which sets out the types of checks required by each type of position.

Supervisor Signature:		**Date:**	10/9/2021
Printed Name:	Seth Lazar	**Uni ID:**	U4925261

References:

Academic Minimum Standards

Pre-Employment Work Environment Report

Position Details

College/Div/Centre	College of Arts and Social Sciences	**Dept/School/Section**	School of Philosophy
Position Title	Research Fellow	**Classification**	Level B
Position No.	TBC	**Reference No.**	

In accordance with the Work Health and Safety Act 2011 (Cth) the University has a duty to provide a safe workplace.

- This form must be completed by the Supervisor of the advertised position and forwarded with the job requisition to Recruitment and Appointments Branch, Human Resources Division. Without this form jobs cannot be advertised.
- This form is used to advise potential applicants of work environment issues prior to application.
- Once an applicant has been selected for the position consideration should be given to their inclusion on the University's Health Surveillance Program where appropriate – see Health Surveillance Procedure
- Enrolment on relevant Work, Health and Safety (WHS) training courses should also be arranged – see WHS Training & Induction
- Consideration should be given as to whether 'Regular' hazards identified below should be listed as 'Essential' in the Selection Criteria

Potential Hazards

- Please indicate whether the duties associated with appointment will result in exposure to any of the following potential hazards, either as a **regular** or **occasional** part of the duties.

TASK	regular	occasional	TASK	regular	occasional
key boarding	☒	☐	laboratory work	☐	☐
lifting, manual handling	☐	☐	work at heights	☐	☐
repetitive manual tasks	☐	☐	work in confined spaces	☐	☐
catering / food preparation	☐	☐	noise / vibration	☐	☐
fieldwork & travel	☐	☐	electricity	☐	☐
driving a vehicle	☐	☐			
NON-IONIZING RADIATION			**IONIZING RADIATION**		
solar	☐	☐	gamma, x-rays	☐	☐
ultraviolet	☐	☐	beta particles	☐	☐
infra red	☐	☐	nuclear particles	☐	☐
laser	☐	☐			
radio frequency	☐	☐			
CHEMICALS			**BIOLOGICAL MATERIALS**		
hazardous substances	☐	☐	microbiological materials	☐	☐
allergens	☐	☐	potential biological allergens	☐	☐
cytotoxics	☐	☐	laboratory animals or insects	☐	☐
mutagens/teratogens/ carcinogens	☐	☐	clinical specimens, including blood	☐	☐
pesticides / herbicides	☐	☐	genetically-manipulated specimens	☐	☐
			immunisations	☐	☐

OTHER POTENTIAL HAZARDS (please specify):

Supervisor's Signature:		**Print Name:**	**Seth Lazar**	**Date:**	**10/9/2021**