Hertie School

Master Thesis Supervision Plan
Academic Year 2020-2021
MIA/MPP

Name:	Joanna J. Bryson
Professor of...	Ethics and Technology

Advisor Information

Office	3.20.1
E-Mail	bryson@hertie-school.org
Faculty Assistant	Alex Karras
Office Hours	To be announced online (weekly but irregular)

Research & Advising Profile:
General Academic Field: What is your academic field, and what are your areas of supervision?

I hold four formal academic degrees: in behavioural science (non-clinical psychology), artificial intelligence (AI), psychology, and computer science (PhD, focus on the systems engineering of AI). I have also published in behavioural ecology, biological anthropology, cognitive science, and philosophy as well as in technology policy, and have a paper near publication in political economy. My fundamental interests are in the nature of intelligence and how it impacts our (and other species!) societies and lives. I came to Hertie School though to focus on better understanding specifically how artificial intelligence alters what we can and should do. I am particularly interested in supervising science-oriented dissertations. These can either:

- contribute to theory construction by modelling the impacts of technologies particularly communication, or (even better)
- contribute to theory testing by finding and analysing and/or produce quantitative evidence confirming, disproving, or otherwise addressing existing theories and models.

Specific Expertise: What is your specific area of expertise, and what topics are you supervising?

I am open to suggested topics, though see above for the types of suggestions I'm most interested in, and also my publications since 2015. In particular, I do not think it is appropriate to do a purely philosophical dissertation in a governance school. Though philosophy can influence the choice of topics and reasoning is central to all academic endeavours, I am interested first, in helping you learn to form policy and/or persuade through the sound use of evidence. A few possible topics:

- Deriving from my 2017 Science paper on word embeddings: 1) What is "unpleasant" about African American names? 2) Can we show whether Whorf is right or wrong? Do labour terms (for example) have different meanings in different languages?
- Deriving from my arxiv paper (hopefully soon published) on political polarisation: 3) who has been able politically to use science to change culture during the pandemic crisis? 4) does the extent of affective bias correlate with the extent of political polarisation?
- Concerning the future of work (see the arxiv / soon to be published paper by David Autor, not me for methodological hints): 5) Why are White men without college educations withdrawing from the workforce in the USA? Is there a reason for why Black men have participated less relative to Black women other than racism, and are White men finding themselves increasingly in the same context?
- Concerning transparency (scientific communication to policy makers or the general public): 6) transparency in agent-based models of the pandemic (or other agent-based or scientific models), 7) transparency in the moral status of human-like AI systems.
- Concerning transparency for accountability and governance 8) transparency in software development by and/or for governments, corporations, or other large agencies, 9) transparency in automated decision making.

Methodology: What methodologies and methods are you able to supervise?

See "supervisory style" below. I am willing to supervise dissertations based on methods I am not myself able to explicitly teach. I can still guide and advise. Interdisciplinary work requires drawing together knowledge from a variety of fields. You may need to take courses in data science or causality for some of the above topics. I am willing to provide moderate financial support for e.g. surveys, data access etc. for good projects. See my publications for examples of methods I have already published on.

Supervision Style: Please be informed that there will be both onsite/online options according to your availability:

I expect masters students to be very capable of self-learning, but also to need guidance in matters of experience such as what are good topics, where can resources be found, how large of a project should be attempted, when during the semester you may want to scale up or down your expectations for the project. I do most of my supervision through weekly group meetings, and encourage students to work together on learning new techniques and discovering resources. I expect students to require one or two individual meetings a semester as well, in order to pick apart very individual problems to their topic. One individual meeting is typically the week before semester or the first week of semester. Because of time pressure, I do not provide much email

support, but would consider setting up slack or teams for the students. I often publish with masters students (and sometimes undergraduates.)

Examples of Previously Supervised Thesis Topics: (*if applicable*)

This is my first year supervising masters dissertations at Hertie School. I was not able to supervise for the past five years at my previous university (Bath) because I was living in the USA, but you can see older dissertations (including PhD and undergraduate) here: http://www.cs.bath.ac.uk/~jjb/here/dissertations.html The PhD dissertations are probably more relevant in terms of topics than those of the masters and undergraduate students studying for computer science degrees, but you may find some interesting topics in those as well.

Recommended General Readings: (*if applicable*)

The following texts are general readings material relevant for all advisees:

See the papers here https://www.joannajbryson.org/ethics-and-policy-of-technology and in particular https://www.bbvaopenmind.com/en/articles/the-past-decade-and-future-of-ais-impact-on-society/

The arxiv paper on polarisation mentioned earlier is here: https://arxiv.org/abs/1807.11477

Proposed Projects: (*if applicable*)

Project Title	Practice Partner (if applicable)
I put these above. There are no partners in particular at the moment, but there could easily be opportunities to collaborate with governments, corporations, or other academic institutions.	

Plagiarism:

Plagiarism is an infringement of § 11 Good Academic Conduct, 2a: "Infringements of the standards of good academic conduct include for instance to use wordings, ideas or other intellectual work of others in an academic work without clearly indicating the author. The obligation to indicate the authorship of others shall apply irrespective of whether or not the sources used are protected by copyright" (See: Exam Rules, § 11 Good Academic Conduct for more information).

It is vital to keep track of your sources and to cite all material properly.

The Library will offer a session available to all students on resource management and proper citation.

Extra (individual) note on plagiarism

Attendance:

Students receive 8 ECTS for the Master´s thesis colloquium. Attendance for supervision—whether in the form of a colloquium or an individual meeting—is compulsory. Students are expected to be present, prepared, and engaged in each session, and to adhere to deadlines set for assignments.

In case of an excused absence, students must notify their supervisor. Absence must be compensated with an additional assignment.

Students on an academic exchange programme in the Fall Semester should be given the opportunity of remote supervision.

Participation, Milestones, Colloquium & Meeting Information:

There will be weekly group meetings at a time that works for all concerned. There will in fact probably be two different group meetings per week to ensure full participation by all students. There will be no formal assignments, but feedback will be given on formative submissions in the order they are submitted.

	Dates	Milestone
Milestones (submissions for formative feedback)	31 August – 7 September	Agree supervision, informal selection of topic, agree group meeting times and membership.
	By 28 September	**One-page proposal** of topic, methods, data sources, etc.
	By 27 November	**Literature and technology review.** Includes at least one chapter of formal writing, plus a full project plan with research and writing milestones. Should include best, worst, and expected case outcomes.
	By 28 February	**Enough research should be done such that a passing dissertation could be written.** Analysis and writing increasingly become the focus of dissertation activity.
	From 29 March through 16 April	**Draft dissertations submitted**, will be read in order of submission, those submitted earlier will probably get more feedback. **Note: there will be no last-minute reading of draft dissertations. There will only be one draft read per student.**