

Practical Methods in Human-Centred AI

ACS P342 / Part II unit - Alan Blackwell & Advait Sarkar

Overview

- **Practical experimental course**
 - lectures provide overview and sample of current research
- **This introduction**
 - general principles, research approaches, strategic trends
- **Specialist lectures**
 - six deep-dive topics, each illustrating some practical methods
- **Design and run your own study**
 - practical feedback on work in progress every week
- **Final presentation of your results**

Course objective

□ “Human-Centered AI” Ben Shneiderman (OUP 2022)

- 1) Process: HCAI builds on user experience design methods of user observation, stakeholder engagement, usability testing, iterative refinement, and continuing evaluation of human performance in use of systems that employ AI and machine learning.
- 2) Product: HCAI systems are designed to be supertools which amplify, augment, empower, and enhance human performance. They emphasize human control, while embedding high levels of automation by way of AI and machine learning. Examples include digital cameras and navigation systems, which give humans control yet have many automated features.

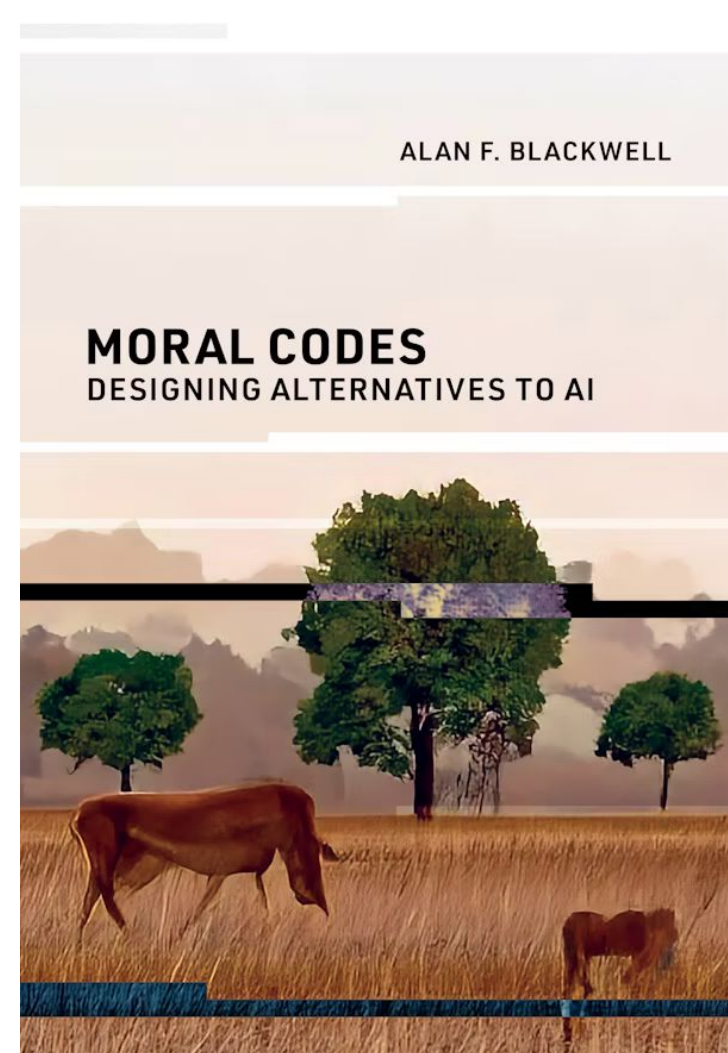


The book of the course (in Alan's mind – Advait has views!)

Moral Codes

Designing alternatives to AI

Alan Blackwell
MIT Press 2024



Where Ben, Alan and Advait agree with a Nobel prizewinner

- Four waves of AI, according to DeepMind founder (& CS grad) Demis Hassabis:
 - First wave (GOFAI): Expert systems & symbolic reasoning
 - Second wave: Statistical inference
 - Third wave: Deep learning
 - Fourth wave: Intelligent tools
- Our approach:
 - Intelligent tools as advanced HCI
 - Including: Programming (agentic AI?), Labelling, Explanation, Authoring text
- A *practical* HCI course:
 - Project work to build, customise, measure, observe ...
- For: Part III and MPhil ACS (research preparation), Part II (advanced HCI)

Your background

- 1. Prior HCI experience
- 2. Prior ML/AI experience
- 3. What do you hope to get out of this course?

	None	Casual	Student	Professional
HCI			8	6
ML			13	1

Target outcome

- This is a specialised and focused practical research training course.
- The expected outcome:
 - You will achieve research competence in a recognised academic field such as Intelligent User Interfaces, Interactive Intelligent Systems etc
- ACS assessment will be relative to the international standard of graduate students working in these fields.
 - Written work will be graded relative to typical student publications in the field
 - Presentations will be expected to meet the standard of first-year PhD students in the field, for example at the Doctoral Consortium of a specialised conference.
- Part II students demonstrate skills by “replicating” a competent study.

Lecture topics

- **Week 2 - Mixed initiative interaction (AB)**
 - information gain, cognitive ergonomics, agency & control
- **Week 3 – Labelling as a fundamental problem (AS)**
 - attribution, subjectivity, reliability, consistency
- **Week 4 - Program synthesis (AB)**
 - end-user programming (= agentic AI?), attention investment
- **Week 5 - Generative AI (AS)**
 - creativity and knowledge work
- **Week 6 – Bias and fairness (AB)**
 - discrimination, accountability and ethics in hybrid systems
- **Week 7 - Explainability (guest)**
- **Week 8 – Your research presentations**

Practical work plan

- Week 1 - select research question
- Week 2 - discuss potential study approaches
- Week 3 - review and feedback on study proposals
- Week 4 & 5 - review logistical issues / practical progress
- Week 6 - discuss preliminary findings
- Week 7 - discuss research implications
- Week 8 - final presentation

Assessment for ACS

- Final research report (80%)
 - Based on your practical work
 - Presented as an original research paper
- Optional (but recommended) work-in-progress drafts
 - Advisory grades will be provided as feedback, for revision in final report
- Reflective diary (20%)
 - Summarise lectures
 - Document discussions
 - Record development of your own thinking
 - Make 8 weekly entries ...
 - ... bind together and submit with a final summative review

Assessment for Part II

- **Final research report (80%)**
 - Based on your practical work
 - Presented as a research paper replicating a previous publication
- **Ticks awarded for work-in-progress drafts (20%)**
 - Advisory grades will be provided as feedback, for revision in final report

Practical work-in-progress

- Week 2 - Research question (200 words) + a sample diary entry for ACS
- Week 3 - Study design (400 words)
- Week 4 - Another sample diary entry for ACS
- Week 5 - Draft literature review for final report (400 words)
- Week 6 - Draft introduction to report (200 words)
- Week 7 - Draft results section for report (400 words)
- Week 8 - Draft discussion section for report (200 words)

“Indicative feedback” on work in progress

- A+ excellent - on target for 85-100
 - A very good - on target for 75-85
 - B good - on target for 70-80
 - C acceptable - on target for 60-70
 - D disappointing - risk of fail
-
- The final grade will be awarded solely on the basis of the final report, and you are welcome to change as much as you like in response to feedback, or to simply copy draft material straight in, whichever you prefer.



Theories of interaction

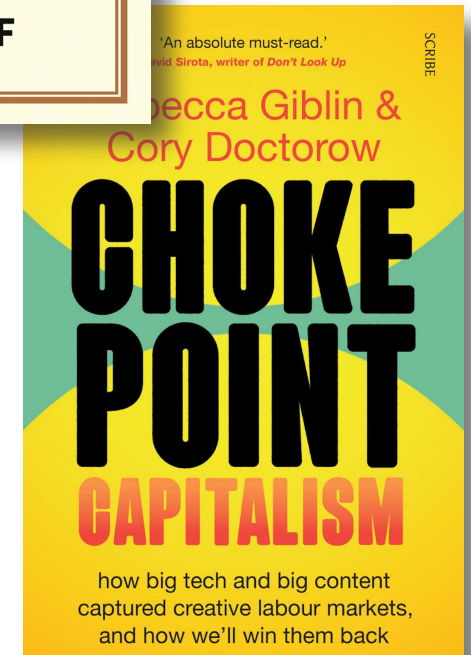
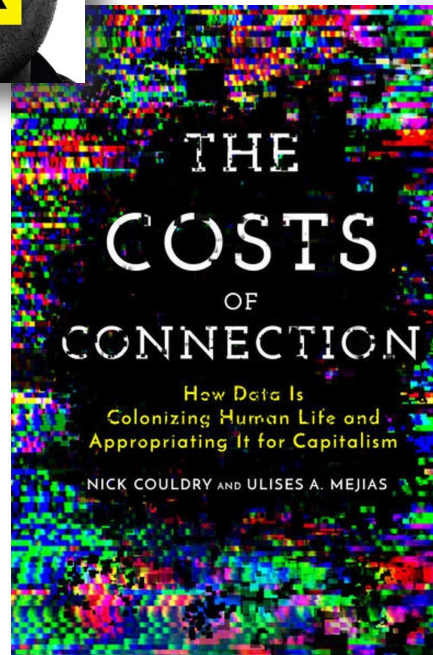
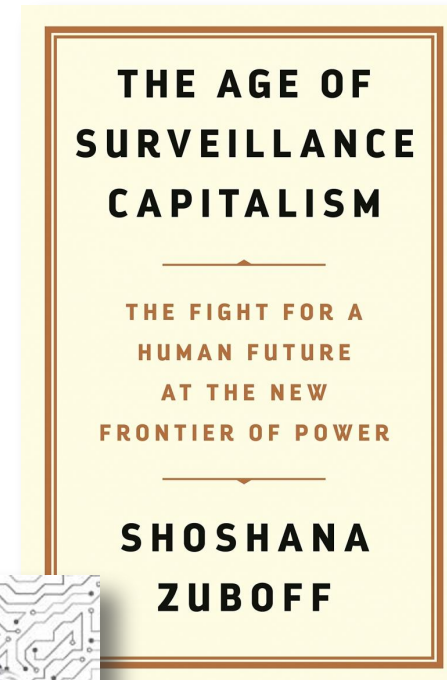
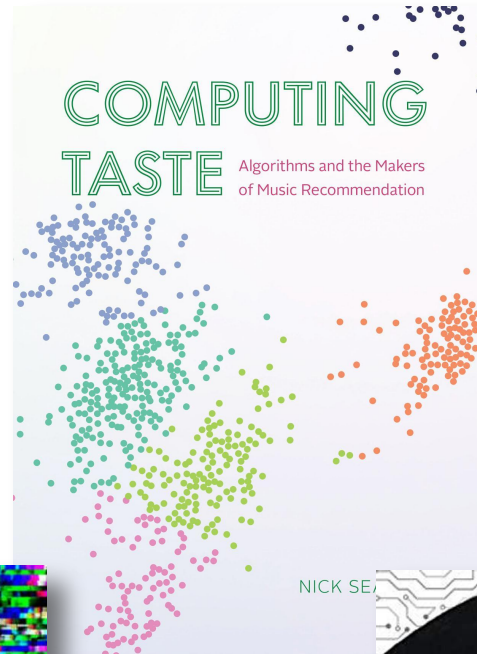
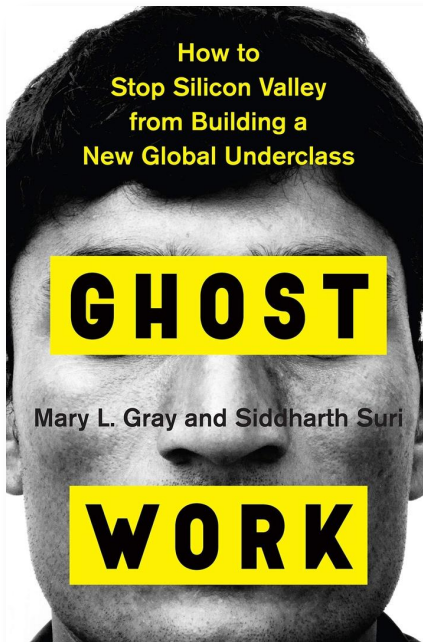
Human-Computer Interaction (HCI)

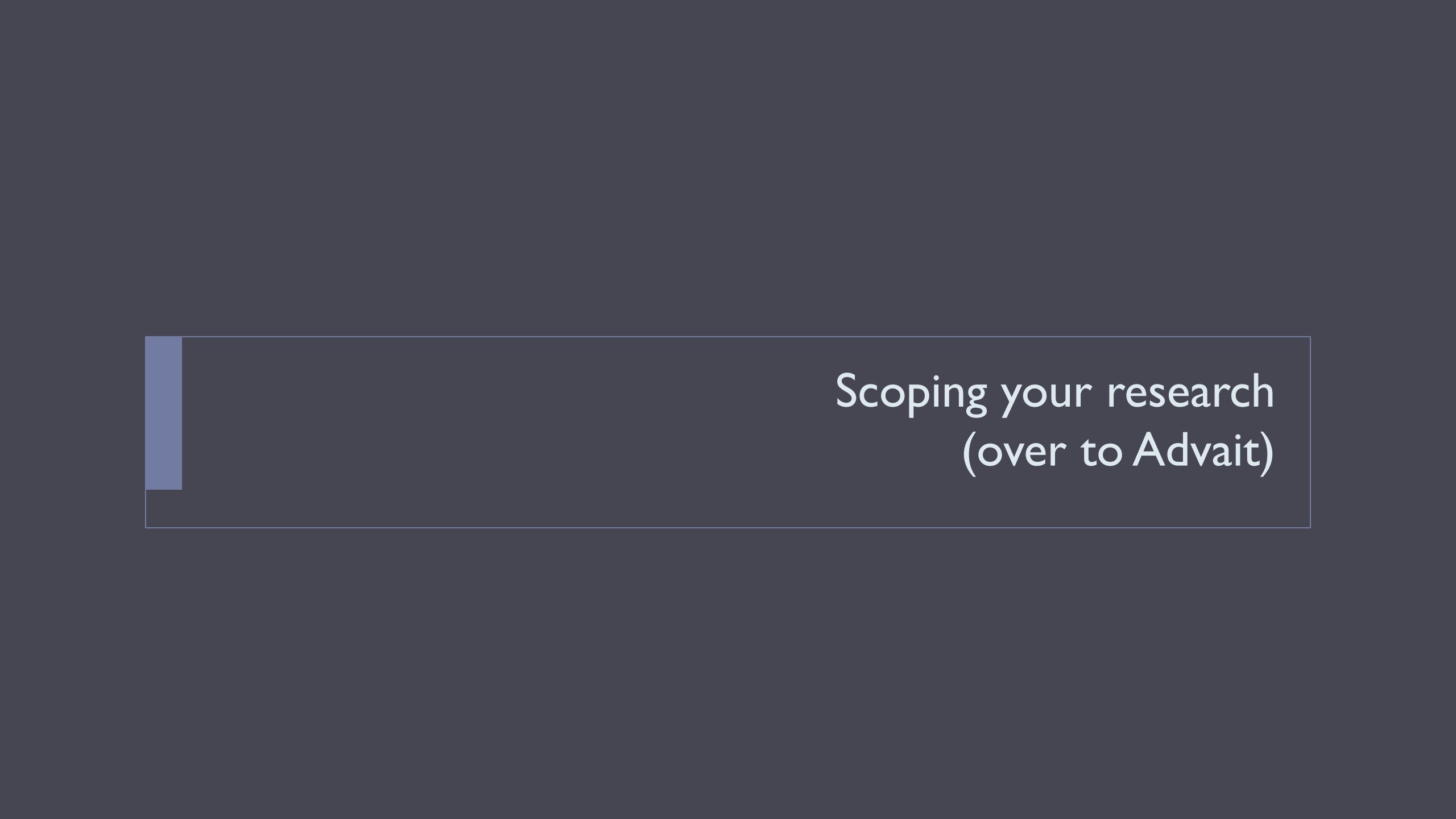
- First wave (1980s):
 - Theory from Human Factors, Ergonomics and Cognitive Science
- Second wave (1990s):
 - Theory from Anthropology, Sociology and Work Psychology
- Third wave (2000s):
 - Theory from Art, Philosophy and Design
-
- Text books:
 - Preece, Rogers and Sharp (2019) *Interaction Design beyond HCI*
 - Cairns and Cox (2008) *Research Methods for Human-Computer Interaction*
 - Carroll (2003) *HCI Models, Theories and Frameworks*

Older paradigms of intelligent user interfaces

- Perfect information games (toy worlds, chess, go, videogames)
 - Not considered particularly interesting
- Recommender systems
 - Once a major research area, now familiar - Amazon, Spotify, YouTube, Netflix, etc.
- Scripted dialogue / heuristic-based chatbots and agents
 - e.g. voice assistants – but watch “guardrails” become recommenders!
- Programming by example, program synthesis (recently agentic)
 - See Lieberman *Watch What I Do*, but also e.g. Microsoft Excel FlashFill
- Human-in-the-loop automation
 - Autopilots, remote-operation, “autonomous” vehicles
- Stochastic algorithms as creative aids
 - Art, creative writing, music, dance

AI, Ethics & Society - topics considered in R225 last year





Scoping your research
(over to Advait)