

**AI Ethics Resources, compiled by Tommy Kessler**  
(to be integrated into the course as Learning Modules, TBD)

**ETHICS**

Virtue Ethics Preliminaries: <https://plato.stanford.edu/entries/ethics-virtue/#Prel> (Section 1.1 through 2.1)

What is the importance of practical wisdom?  
Is there such a thing as “true” (objective) happiness?

Kant: <https://www.thoughtco.com/kantian-ethics-moral-philosophy-immanuel-kant-4045398>

Can morality exist outside of a religious context?  
Is moral obligation a solid enough base upon which to base one’s conception of the “good”?

Social Contract Theory:

[http://www.sophia-project.org/uploads/1/3/9/5/13955288/elahi\\_socialcontract.pdf](http://www.sophia-project.org/uploads/1/3/9/5/13955288/elahi_socialcontract.pdf)

Do we need government?  
What amount of power can our leaders have and still be effective? Is a “sovereign” necessary?

**AI ETHICS**

AI/Machine Learning Crash Course: [https://www.youtube.com/watch?v=f\\_uwKZIAeM0](https://www.youtube.com/watch?v=f_uwKZIAeM0)

Intro to Ethics of Technology: <https://plato.stanford.edu/entries/technology/#DeveEthiTech>  
Read: 3.1 & 3.3

Al Jazeera Video AI Ethics Overview: <https://www.youtube.com/watch?v=iyG-A09bBt0>

Joanna Bryson: HAL (not the 2001 machine!): <http://www.cs.bath.ac.uk/~jjb/ftp/HAL00.html>

Why is it necessary to set limits on the uses of a technology before it is sold?  
Is it important to put the needs of a society over the needs of an individual? How does this play into establishing an ethics of AI?

Joanna Bryson: Robots Should Be Slaves:

<http://www.cs.bath.ac.uk/~jjb/ftp/Bryson-Slaves-Book09.pdf>

Who should be responsible for a robot’s actions?  
What is the role of programming in regard to AI ethics?  
Why do you think that people wish to regard robots as peers?

Kate Crawford Interview:

<https://www.wired.com/story/why-ai-is-still-waiting-for-its-ethics-transplant/>

Why is it important AI research be cross-disciplinary?  
Why can't data be neutral/neutralized?

## **BIAS**

Why isn't AI decision-making inherently objective?  
What can we do to help combat bias in AI?  
What are some reasons that facial recognition is dangerous?

Racial Bias in AI Decision Making:

<https://fivethirtyeight.com/features/technology-is-biased-too-how-do-we-fix-it/>

Gender Bias:

<https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>

Sexual Orientation:

<https://www.theverge.com/2017/9/21/16332760/ai-sexuality-gaydar-photo-physiognomy>

## **TRANSPARENCY**

<https://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial>

Why is it necessary or unnecessary to understand how an AI operates?  
Whose job is it to set rules for transparency?

<https://www.wired.com/story/our-machines-now-have-knowledge-we-never-understand/>

What does our reliance on machine models for knowledge say about our trust of machines? Is this trust warranted?

What have machines revealed to us about our previous models of the universe? What does this mean for science?

<http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable>

What are some of the issues in defining "interpretability"?  
What are "suitcase words" and how do they relate to interpretability?

## **MISC**

Automation & Employment:

<https://www.technologyreview.com/s/603465/the-relentless-pace-of-automation/>

Should we sacrifice human employment for the sake of efficiency and technological advancement?

What measures can be put in place to proactively help those displaced by technological advancements?

Social Media: [https://motherboard.vice.com/en\\_us/article/kbzmvm/weapons-of-mass-persuasion](https://motherboard.vice.com/en_us/article/kbzmvm/weapons-of-mass-persuasion)  
Should platforms ban the spreading of false or misleading information?

Spreadsheet of Tech/CS Ethics Curricula:

<https://docs.google.com/spreadsheets/d/1jWlrA8jHz5fYAW4h9CkUD8gKS5V98PDJDymRf8d9vKI/edit#gid=0>

Harvard Course: The Ethics and Governance of Artificial Intelligence by Griswold505

<https://h2o.law.harvard.edu/playlists/53282>