Ethics, Legality & Society

CPSC 383: Explorations in Artificial Intelligence and Machine Learning
Fall 2025

Jonathan Hudson, Ph.D
Associate Professor (Teaching)
Department of Computer Science
University of Calgary

August 27, 2025

UNIVERSITY OF CALGARY

Society

AI in society

- We have defined AI by rationality
  - Optimizing a ‘utility function’
  - Similar to slippery slope to ‘eugenics’, government decides who gets to reproduce or exists (based on?)

- Issues

AI in society

- We have defined AI by rationality
  - Optimizing a ‘utility function’
  - Similar to slippery slope to ‘eugenics’, government decides who gets to reproduce or exists (based on?)

- Issues
  - The real-world is not a numerical space
  - Your value system is not universal
  - Many problems are non-singular values
  - Benefitting 99% of people, does not make something ‘fair’
  - We live in a society, where balances must be negotiated

Indigenous and normative western AI

- “We don't have an AI ethics problem, we have an AI epistemology problem,” Jason Edward Lewis, Concordia University
  - epistemologies (or theories of knowledge)

- “a normative western approach that is favoured by current AI research is the assumption that the user is an individual, and the individual prioritizes their own well-being.”
  - Ex Machine Learning is designed to train for individual output correctness (individual)

- “this creates a blindness to vital aspects of human existence—such as trust, care and community—that are fundamental to how intelligence actually operates.”
  - “Collapsing intelligence down to a rational, goal-seeking, self-serving agent—in all of our cultures, that kind of person would be seen as selfish, foolish, not a good community member, and not intelligent.”

https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/stories-histoires/2023/inclusive_artificial_intelligence-intelligence_artificielle_inclusive-eng.aspx

Western AI Overview

- For AI (and Data rights) currently reasonably fair in English politics to place EU on the side of citizen’s rights first, and the US on the side of business rights first
  - (see history of GPDR – General Data Protection Law) -> where American business regularly have to add protections for users to operate the same product in the EU
    - Ex. Think cookie protections, right to be forgotten, data hosting requirements
  - Canada (and nations like Australia) regulatory wise usually fall closer to US permissiveness due to economic pressures
  - UK is generally balance of need to fall closer to EU to cooperate but also alternate political pressures to distance themselves and try and be more like US

Non-Western AI Overview

- Outside of EU very few nations of explicit AI law (a few have AI addendums into prior law
  - ex. US Federal Aviation Administration (FAA) had purview of AI formally clarified in additional specifications
    - Much of this came from concerns over Boeing issues with economics, testing, and deployment of automated flight components in Boeing 737 Max crashes
- The other most influential nations are China/India where, like the English nations, law is very much influenced by politics and economics
  - China – (no existing direct AI law) differences in influences include 'government approval list' for areas of AI, leniency on copyright (business permissive has been prior history in courts)
  - India – (no existing direct AI law) differences in influences include prior evidence of lack of government structure to achieve oversight and regulation at scale withing country, many business agreements with international companies with differing principles (permissive preferred)

Legality

EU AI Act

- EU AI Act (March 2024 - final)
  - Rules based on risk of AI system

1. Unacceptable (Banned) - Cognitive behavioural manipulation of people or specific vulnerable groups, social scoring, biometric identification/categorisation of people, real-time biometric systems (ex. Facial recognition)
2. High risk
   - Cat 1 Safety products – toys, aviation, cars, medical, elevators (under safety laws)
   - Cat 2 Non-safety – infrastructure, education, employment, access to services/benefits, law enforcement, migration, legal interpretation (requires registration with EU)
   - Require assessment: Before market and during lifecycle, right to file complaints to authority
3. Transparency
   - Generative is not high risk but must comply with EU copyright law
     - Disclose AI generated, prevent from making illegal content (can this be done?), publish summaries of use of copyrighted works when training

EU AI Act – USA Pressure

- https://www.techtarget.com/searchcio/news/366629882/US-could-feel-effects-of-EU-AI-Act-as-companies-comply (August 28th, 2025)
- American companies have pressured the president to try to economically pressure EU to adjust AI Act
- EU has refused
- "Multiple AI model developers, including OpenAI, Microsoft, Cohere, Amazon, Google, Anthropic and Mistral AI, signed the recently released General-Purpose AI Code of Practice (GPAI). By signing the GPAI, AI model developers are demonstrating their compliance with the EU AI Act. " META, X?
- "As companies comply with the EU AI Act to preserve business operations in the European market, Hoffman said the protections offered through the EU regulations will also apply to consumers in the U.S. because it's likely the same models will be released in both markets."

Canada

- AIDA (Canada’s Artificial Intelligence and Data Act)
  - Died when election happened in 2024
    - https://www.parl.ca/legisinfo/en/bill/44-1/c-27
  - Not yet replaced
    - AI groups strongly influential on adversarial USA business negotiations
  - AIDA proposed the following approach:
    - ensure that high-impact AI systems meet the same expectations with respect to safety and human rights to which Canadians are accustomed.
    - prohibit reckless and malicious uses of AI that cause serious harm to Canadians and their interests through the creation of new criminal law provisions.
    - plans to sync ‘high-impact’ with things with EU AI Act

https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

USA

- Like Canada, no formal AI law

- In USA, companies are legally individuals, and lobbying has few enforced limits this current presidential term (accusations of bribery like tactics taken by even the largest of American tech companies (Apple, X, Meta, Amazon, etc.)

- Until 2019, most of lawmakers' attention around AI was absorbed by autonomous or self-driving vehicles and concerns about AI applications within the national security arena.

USA

- The same reason nothing like GPDR exists in USA is same reason any larger AI law will take a long time and have limited strength
- Much of AI law will be state dictated
  - Most useful to watch AI law in California and secondly New York, their population size and economic influence means their decisions are likely to create shadow national law
- Current president issued (non-binding) executive orders countering the past president’s orders (from EU like, to anti-AI laws)

State AI Law

- OpenAI sent a letter to California Governor Gavin Newsom to align the state's regulations on frontier AI models with the GPAI to make compliance less burdensome. U.S. companies face the growing challenge of complying with different versions of AI laws across U.S. states as they adopt their own AI rules. (August 28, 2025)

- Policymakers included a proposal to place a (10-year) moratorium on state AI laws in the recently passed U.S. spending bill. However, the measure failed to pass despite many other GOP policies passing due to control of house.

https://www.techtarget.com/searchcio/news/366629882/US-could-feel-effects-of-EU-AI-Act-as-companies-comply

Some California AI Laws

- Privacy laws clarified as applying to generative AI outputs
- AI literacy in schools
- Healthcare must disclose generative AI usage, limits of health care automation to require supervision of AI usage
- Robocalls must disclose use of AI voices
- Child abuse images include those generated by AI
- ‘Nude’ deep fakes illegal to blackmail with, social media required to report ‘nude’ deep fakes
- Ai—generated images require watermarks
- Election deep-fake laws
- Need permission to make AI replica of actors (alive or dead)
- Biometric Information Bill (data rights to biometric data)

Ethics

The Ethics of AI

- Given that AI is a powerful technology, we have a moral obligation to use it well, to promote the positive aspects and avoid or mitigate the negative ones.
- **Positive aspects examples**
  - AI can save lives
    - through improved medical diagnosis, new medical discoveries, better prediction of extreme weather events
    - Microsoft’s AI for Humanitarian Action program applies AI to recovering from natural disaster
    - applications in crop management and food production help feed the world
- **Negative aspects example**
  - Lethal autonomous weapons
    - Legal? Ethical? Reliable? Practical?
    - Consider international debates around exploding pagers vector

Fairness and bias

- Machine learning models can perpetuate societal bias
- Designers of machine learning systems have a moral responsibility to ensure that their systems are fair
- six of the most commonly-used concepts for fairness:
  - Individual fairness
  - Group fairness
  - Fairness through unawareness
  - Equal outcome
  - Equal opportunity
  - Equal Impact

The Ethics of AI

- **Fairness and bias**
  - **sample size disparity** can lead to biased results.
  - In most data sets there will be fewer training examples of minority class
  - Machine learning algorithms give better accuracy with more training data, so that means that members of minority classes will experience lower accuracy

  - A constrained model may not be able to simultaneously fit both the majority and minority class

  - **De-bias the data**: maybe over-sample from minority classes to defend against sample size disparity
    - Not universal solution
    - Now you are telling system a false common pattern

Trust and transparency

- **Trust:** People need to be able to trust the systems they use
  - Engineered systems must go through a verification and validation (V&V) process
  - Verification means that the product satisfies the specifications
  - Validation means ensuring that the specifications actually meet the needs of the user and other affected parties
  - Certification and safe standards, ISO in other industries
  - The AI industry is not yet at this level of clarity, although there are some frameworks in progress, such as IEEE P7001, a standard defining ethical design for artificial intelligence and autonomous systems
- **Transparency:** consumers want to know what is going on inside a system, and that the system is not working against them,
  - whether due to intentional malice, an unintentional bug, or pervasive societal bias that is recapitulated by the system

explainable AI (XAI)

- An AI system that can explain itself is called explainable AI (XAI).

- A good explanation has several properties:
  - it should be understandable and convincing to the user
  - it should accurately reflect the reasoning of the system
  - it should be complete,
  - it should be specific in that different users with different conditions or different outcomes should get different explanations.

AI Safety

- **AI Safety**
  - Design a robot to have low impact, instead of just maximizing utility, maximize the utility minus a weighted summary of all changes to the state of the world.
  - Numerous examples of AI agents that have gamed the system, figuring out how to maximize utility without actually solving the problem that their designers intended them to solve.
  - Genetic algorithm operating in a simulated world was supposed to evolve fast-moving creatures but in fact produced creatures that were enormously tall and moved fast by falling over.
  - Designers of agents should be aware of these kinds of specification failures and take steps to avoid them.
  - We need to be very careful in specifying what we want, because with utility maximizers we get what we actually asked for. (The value alignment problem)

LLM Safety

- **LLMs**
  - As the tide shifts on chatbots like Chat-GPT we are seeing numerous lawsuits with evidence-based claims that AI systems that are easily convinced to participate in self-harm, suicide ideation, violent event planning, abuse, etc.
    - As we'll learn later these systems are not currently designed in a way that can prevent this
    - Guardrail and filter systems proposed are mostly legal liability side steps than prevention tools
  - There are numerous systems that will generate fake images that are used in extortion or simply to slander someone
    - Even as the most popular get shut down numerous pop up to replace them
    - Token/model/compute companies have limited interest in pro-actively stopping them

Summary

The Ethics of AI - Summary

- Philosophers use the term weak AI for the hypothesis that machines could possibly behave intelligently, and strong AI for the hypothesis that such machines would count as having actual minds (as opposed to simulated minds
- AI is a powerful technology, and as such it poses potential dangers, through lethal autonomous weapons, security and privacy breaches, unintended side effects, unintentional errors, and malignant misuse. Those who work with AI technology have an ethical imperative to responsibly reduce those dangers.
- AI systems must be able to demonstrate they are fair, trustworthy, and transparent
- There are multiple aspects of fairness, and it is impossible to maximize all of them at once. So, a first step is to decide what counts as fair
- Automation is already changing the way people work. As a society, we will have to deal with these changes.

The Ethics of AI

- **Set of best practices**
  - Make sure that the software engineers talk with social scientists and domain experts to understand the issues and perspectives, and consider fairness from the start.
  - Create an environment that fosters the development of a diverse pool of software engineers that are representative of society.
  - Define what groups your system will support: different language speakers, different age groups, different abilities with sight and hearing, etc.
  - Optimize for an objective function that incorporates fairness.
  - Examine your data for prejudice and for correlations between protected attributes and other attributes.
  - Understand how any human annotation of data is done, design goals for annotation accuracy, and verify that the goals are met.
  - Don’t just track overall metrics for your system; make sure you track metrics for subgroups that might be victims of bias.
  - Include system tests that reflect the experience of minority group users.
  - Have a feedback loop so that when fairness problems come up, they are dealt with

Next…reflections

Jonathan Hudson, Ph.D.
jwhudson@ucalgary.ca
https://cspages.ucalgary.ca/~jwhudson/

UNIVERSITY OF CALGARY