MAY 2021

EDUCATION PROGRAM

AI + SDG Launchpad

Overview

AI for Good

FOUNDATION

Introduction

The Sustainable Development Goals are the rallying call of our generation. In an enormously complex and interconnected world, we have watched as some of our most fragile ecological structures have been driven to extremes; we have seen only marginal improvements in the delivery of basic literacy and numeracy; universal healthcare is a far-off aspiration; and half of our population participates in society as second class citizens. The acceleration of technology adoption has levelled some playing fields, and widened others. Access to technology plays a role in concentration of wealth and inequality and puts editorial news outlets on a level playing field with advertising networks and propaganda peddlers.

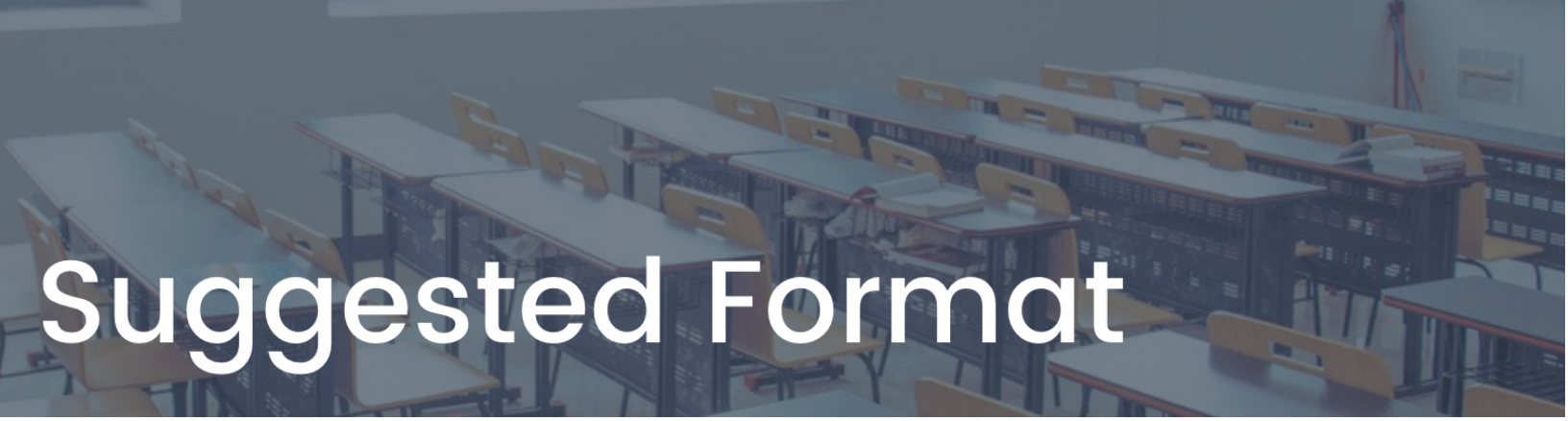
How should technology innovation (and its prodigal child—Artificial Intelligence) be leveraged to support a positive impact on these critical objectives? How do we research, design, debate, plan, and execute interventions that are both human and technological at heart? How do we engender resilient communities with bold plans to make them healthier, happier, and more sustainable?

The purpose of the SDG Launchpad is to act as a foundational curriculum to build the technically sophisticated Policy Scientists of the future—a set of changemakers to operate at all levels of society and around the world to bring the best out of humans and technology for the benefit of society and our planet.

This course proposes a blend of case-based analysis, flipped classrooms, policy development, analytical research, ethical exploration, thought experiments, technical implementation, and a capstone project designed to create real impact. Together, these elements help elucidate some of the biggest frictions and inefficiencies in the policy intervention “marketplace”,

and set up the student to be part of a growing community of ethical stewards of 21st century “innovation-led governance”. After the course, alumni are invited to join our Launchpad community, become mentors, get involved in further projects, attend global leadership summits, and participate in crafting the global debate.

Table of Contents



Suggested Format

The course can be planned over one or two semesters, with two semesters being the preferred approach. In one semester, it is recommended to drop the capstone element and focus on discussion-based classes and a mixture of individual and collaborative assignments. For the two-semester version, all elements may be covered, with an optional summer extension at schools who support summer research opportunities.

We recommend a class size of 25-150 students from our experience, with around 50-60 being the “sweet-spot” for sufficient exposure to the assignments and to other groups’ work, while not drowning in groups and losing the intimate nature of some of the discussions. With larger classes, the assignment structure would need to be amended, and presentation/communication de-emphasised for practicality.

Additionally, due to there being several instances where assignments requiring a larger time commitment occur between the first and second lecture of the week, we recommend a non-standard class schedule of Mondays and Thursdays, or Tuesdays and Fridays—allowing a more even split of homework time between lectures.

The course works best with students who have taken an introductory CS sequence, as well as demonstrated interest in the social sciences. Economics and econometrics students have done especially well, with pure CS and pure Politics students faring consistently as well. It can be beneficial to aim for a blend of policy students, ecology, philosophy, and broader engineering disciplines, as this increases the perspectives and richness of the class discussions.

In terms of grading, we recommend that these courses be either pass/fail, or graded at a group level. Creating a complicated blended grading rubric is at

your own peril, although we certainly see the attraction of incentivising students to not simply be 'dragged-along' by their group. On the other hand, the purpose of this course is to provide students with new frameworks for thinking at the intersection of science and policy, not to rank students. Although we provide basic materials for teaching the course, and can provide access to key resources such as project mentors, stakeholder connections, etc., the course works best when you take the materials you want and create your own narrative. All material is optional, and there are enough relevant ideas here to craft a large number of independently great single-semester courses. It's up to you.

Most importantly, help us to build a lasting community of young people who will think in this way, no matter where they end up—the world will be better off for it.

Resources Provided by AI for Good

Resources Provided:

- Modular Syllabus
- Basic Teaching Materials
- Example Capstone Projects
- External Partner Project Proposal Form
- Template for Initial Research Brief
- Template for Initial Exploration Document
- Template for Project Plan
- Template for Evaluation Plan
- Framework Recommendation for Code, Data, and Collaboration
- Network of Scientific and Policy Mentors
- Network of Stakeholders
- Hosting of Group Final Materials and Dissemination

Coming Soon- August 2021

The AI for Good Foundation will create a digital and physical textbook for this course, with contributions from leading experts in each relevant sub-discipline. In addition, we will provide open courseware with all lectures freely accessible for incorporation into remote learning structures.

Course Overview

This course is designed to be taught over one semester (12 weeks) or two semesters (24 weeks), with two 1-2 hour lectures each week. We plan for 6-10 hours of student time commitment beyond class time for completing assignments and group work.

Topic 1: Introduction to Course and UN SDGs

Lecture 1: Introduction and Course Overview

This class covers the following topics to set the theme for the course:

- Lecture Content
- Assignments and Expectations
- Readings
- Project
- This lecture covers the learning objectives for the course, the prerequisites, and should provide some examples of how Artificial Intelligence has been used within the context of the SDG's already. Instructors can use example projects from the AI for Good Foundation website, presentations, data sets, etc. as auxiliary teaching materials if they do not have preferred alternative examples.
- Assign students into groups of 4-5 randomly. These act as study groups, and meet each week in order to review the material, discuss any open questions, and to present to each other their completed assignments. The purpose is for the groups to ensure that all students complete work to a minimal standard each week, even if they have already been selected to present in a previous class (the classes have

a large student presentation component, where students are randomly selected to present, therefore this provides a moderating influence to not let down peers). Each week, the groups hear each other's presentation, and each student submits one of 'incomplete', 'pass", 'outstanding' feedback on each other student's work. Use 'outstanding' as a strong signal for higher achievers when creating project groups later, and use multiple 'incomplete' as a strong signal that additional interventions are needed immediately.

Assignment

- Find and summarise three ways in which Artificial Intelligence has been used to support the Sustainable Development Goals. Develop 2-3 slides and come to class ready to present these projects to your peers. Include the following, at a minimum:
    - Name of the project
    - Link
    - Who is involved?
    - Who are the stakeholders?
    - Which SDG does it impact?
    - How is Artificial Intelligence used?
    - What is the impact/how is the deployment evaluated?
- Summary (300-500 words)
- Students should record their 5-minute presentation and submit along with written materials.

Lecture 2: Structure of the Sustainable Development Goals

- The first 10-15 minutes of class to be taken up by 2 randomly selected students presenting their homework assignments. We recommend that the selection takes place at the beginning of class, so that everyone comes ready to present and discuss, however, this is something instructors should decide.

- Overview of the 17 Sustainable Development Goals, and the Millennium Goals that preceded them.
    - Structure of the SDGs
    - Sub-structure—measurable impact
    - Examples

Assignment

- Selected Reading: First two chapters of “Artificial Intelligence and International Affairs: Disruption Anticipated” by Jacob Parakilas, Hannah Bryce, Kenneth Cukier, Heather Roff, Missy Cummings.

Topic 2: Research and Science in Society

Lecture 1: Nexus of Policy and Research

- Who funds research?
    - In the US (NSF, NIH, DARPA, IARPA, Private Foundations, Corporations)
    - In Europe (European Commission, State Governments, Private Foundations, Corporations)
- The direction of research is guided by the resources available—if there is a large pot of funding for exploring video recommendation, it is likely that research will take place on video recommendation.
- What are some issues with this?
    - Often the expected results are implicit in the research questions/funding sources. Corporations have clauses in agreements to veto research publication, and private funds have missions to propagate particular messages/solutions, often driven by single individuals, families, or homogenous stakeholder groups. Can you find examples? (Gates Foundation?)
- Case Study: Breastfeeding Research
    - Take students through our Breastfeeding Research Case Study, or identify a different topic with the same underlying message.
    - This topic is purposefully controversial, in order to encourage students (who have likely been exposed to the debate in the media) to explore their stances, and deepen the relationship between media, common wisdom, scientific experimentation, validation, and real-world community pressures. The intention is not to create a conservative platform, or to take a stance on breastfeeding. It is clear that breastfeeding does have advantages for mother and child in some respects, and interacts negatively in other ways (inequality, career success, family pressure, etc.). That's what makes this such an interesting case study every time.
    - Our case study explores the fact that the WHO's recommendations for breastfeeding are primarily driven by a pro-breastfeeding lobby group, and substantially funded by the Family Larsson Rosenquist Foundation. The FLRF-funded WHO

committee has pushed for countries to take steps to ensure all new mothers are provided with resources to encourage breastfeeding (pre- and post-natal lactation consultants). The WHO points to scientific research which has, in turn, been funded by FLRF at various academic labs. All research funded by FLRF concludes that breastfeeding should be universally encouraged. FLRF's main source of funding is from the Larsson family's wealth as owners and founders of Medela, the world's leading breast pump manufacturer. That's a lot of pressure to find the right scientific results.

Assignment

- Choose a topic that you are passionate about, and identify who are the main research groups involved in its exploration? Who is funding this research and what are (plausibly) their expectations of the research results? How much funding comes from different sources? Do the scientific results point to a diversity of effects, or do results align with the agendas of the funding parties?
- Put together a report presentation (a few slides + 300-500 word summary) that answers the questions above, with relevant links to first-party information sources, and your overall conclusion as to how trustworthy the results are from a procedural perspective.

Lecture 2: Evaluating Science in the Public Domain

- Choose two students at random to present their assignment at the start of class.
    - 5 minutes each
    - 5 minutes for instructor-led discussion
- The remainder of this class focuses on comparing claims made in the media with the underlying scientific evidence. The aim is to take students through a couple of real-world examples to show the nuance that exists in interpreting statements of "fact" made even in respectable news outlets with a strong editorial process.
- This is not a class about 'politifact'. These are not the kind of statements that a fact checking service would be able to disentangle, and we are not referring to 'quoted statements', but the assumed truth of the article

author. This is a foundational class before diving deeper into scientific method.

- "AI is stealing jobs"—often a statement or basic assumption of news articles. This is a good one to explore due to the proximity of the topic. The statement depends on what 'stealing jobs' means. Is it true that AI replaces some tasks? Most likely. Is it true that individuals are systematically losing work opportunities and positions due to the deployment of Artificial Intelligence? Not supported by evidence in BLS data or granular employment data. Is it true that AI is causing the number of jobs available in the US economy to shrink, or not keep up with population growth? No. In fact, AI appears to be having the opposite effect to date. For this discussion, it can be useful to start from the "evidence", and ask students whether they think it supports the statements. For instance, this page details a variety of 'facts' about AI and jobs, all of which are projections based on guesses about the potential for AI technologies in different sectors, not on the dynamics of employment or any real-world measurement.
- "The use of masks to slow the spread of COVID-19 is supported by scientific evidence". This is a timely topic, so interesting to cover. It is also fairly straightforward to identify the main studies cited by the WHO, CDC, and other relevant state and local health authorities. Ideally, have students spend 10 minutes identifying the main cited studies and aggregate them across the class. What is the experimental setup? To what extent does the experiment show that masks are effective in preventing the spread of COVID-19, as opposed to other confounds (e.g. mandated lockdowns, reduced travel due to fear, etc.)? Does the study support the use of masks more or less in particular situations (inside, outside, etc.)? How direct is the link between the experiment that was run and real-world scenarios? Does it offer an upper bound or lower bound relative to reality? Should the original claim be strengthened or qualified? Are there negative externalities? On net, is the statement useful?

Assignment

- Find a similar statement to those explored in class, and perform the exercise independently.
    - Map out as many linked claims as possible that you can find in the broad cross section of the media.
    - Link each claim to relevant/cited scientific studies.
    - For each study, do the data, methods, and research results support the claim fully?
    - Should the claim be strengthened or qualified?
    - What is the overall takeaway from doing this exercise?
- Come ready with 2-3 slides and 300-500 word write-up to present your exploration to the class.
- Reading: “Popper” by Bryan Magee. This book explores the evolution of thought around scientific method, falsifiability, and in particular how the work of Karl Popper interacted with this topic and the topic of knowledge in society. At 100 pages, it’s the perfect couple of hours of reading to prepare students for the class discussions.

Topic 3: Indirect Consequences and Adverse Effects

Lecture 1: Interrelation of SDGs

- Indirect Consequences and Adverse Effects
    - First, select two students to present their work
- How are different goals connected?
    - Consider the case study of breastfeeding
        - Goals of Good Health and Nutrition, the WHO, and other organisations, recommend breastfeeding children as long as possible (up to 18 months+) in order to safeguard the health of the child, encourage strong family bonds, and reduce the risk of the development of certain allergies.
        - Goals of Gender Equality and Economic Resilience are affected as women feel pressure to breastfeed, which has unavoidable ramifications on career progression, and increases the likelihood that some women will leave the workforce entirely.
    - Class discussion—are there additional spill-over effects among these goals (such as Poverty?)? How can we reconcile competing pressures?

Assignment

- Select a detailed sub-goal of the SDGs. Analyse its potential impact on other SDG areas and metrics. Identify relevant studies that quantify or evaluate the relationship. Develop 3-4 slides of presentation materials to lead a class discussion around this topic. Submit a 1-2 page brief, which offers summary, analysis, and possible ways to reconcile the issues identified.

Lecture 2: Presentation and Roundtable of Previous Assignment

- In this class 3-4 students are randomly selected to present their assignments and the instructor facilitates class-wide round-table discussions of the topics. At this point ~10 students will have been called on to present. Depending on class length and dynamics, you may want to increase or decrease student presentations accordingly. Going

forward, each week's first lecture introduces a topic, and the second lecture hosts student presentations and a facilitated class discussion.

Assignment

- Readings from "Causality" by Judea Pearl, and "Causal Inference" by Scott Cunningham.

Topic 4: Getting at Causality through Statistical and Econometric Analysis

Lecture 1: Causality

- What does it mean to say that “AI causes jobs to disappear”, or “UBI reduces poverty”, or “smoking causes cancer”, or “leaving the EU will make Britons better off”? Statements that imply a causal connection between two variables are necessary for humans to be able to effectively reason about the world, form beliefs, and make consistent decisions. However, these statements should not be taken lightly, since they can lead us to make sweeping policy decisions that might otherwise not have the desired effect, or to have a much smaller effect than we were anticipating.
- Cover key concepts from the readings for the week. Instructor-led.
- Methods to establish a strong likelihood of a connection between two variables:
    - Randomised Control Trials
    - Plausible Natural Experiments
        - Clean Event Studies
        - Plausibly Exogenous Shocks
        - Instrumental Variables Approaches
        - Given real world data about a phenomenon of interest, what other approaches might you follow in order to convince the scientific community that the world is as you see it?
    - Correlative Studies
        - Observe the strength of co-movement in variable values.
        - Difficult to rule out an alternative explanation.
        - Omitted variable bias: are you capturing the full set of interacting variables, and to what extent are their effects overlapping/reinforcing/mitigating?
        - Exercise: How does the likelihood of a spurious relationship change as the size of your data set grows, or the number of features you are able to measure grows?

Assignment

- Make a causal statement that (a) you reasonably believe to be true, (b) is related to the SDGs. Design a series (2-3) of potential interventions (given access to the relevant population, reasonable cash, computation, and storage resources) in order to validate the causal relationship between your chosen variables. Keep in mind the items covered in class, and produce a report (2-3 pages) and presentation that describe the causal relationship, summarise different views around the relationship, describe in detail the interventions, how you would go about determining that the causal relationship has been sufficiently established, and an overall conclusion. Come to class prepared to present.

Lecture 2: Presentation of Reports

- Randomly select 4-5 students to present their assignment for 5-10 minutes each, with instructor-facilitated discussion.

Assignment

- For a Sustainable Development Goal of your choice, research as many openly available data sets as you can find online or through other research portals available to you through university resources and subscriptions. For each data set, identify the following information:
    - Who owns the data?
    - In what publications/studies has it been used?
    - Is it openly available for research purposes?
    - How many samples are included?
    - How does coverage compare to the breadth of the issue in the real world?
    - What attributes are included in the data set?
    - Are there key attributes that are missing?
    - Are the data detailed enough to be used as a basis for scientific research, or are they simply summary statistics of someone else's research?
- Every SDG is associated with hundreds of openly available data sets. Aim to identify at least 10. The more the better! To start, look at AI for Good's SDG datasets.

Topic 5: AI and the Sustainable Development Goals

Lecture 1: AI and the Sustainable Development Goals

- Randomly select 2 students to present their assignment for 5 minutes each.
- This week is all about relating the Sustainable Development Goals to potential Artificial Intelligence-enabled solutions and supporting infrastructure, while keeping in mind all of the areas that have been explored to date in the class.
- Case Study: Analysing a Potential Question (e.g., “Ocean Plastics”)
    - What are the key issues?
        - Why is it a problem?
        - Who/what causes the problem? Why? What is the friction?
        - Who are the main stakeholders with legislative or administrative responsibility, social capital, or other incentives to act?
    - What logical reasoning or inference needs to be scaled?
    - What data sets exist or need to be created that relate to the existing issues?
        - How do they relate?
        - How complete are they?
        - How reliable are they?
        - Are the data collected fairly, ethically, and consistently?
        - How much history is available?
        - If the same information were collected through alternative channels, would the data look the same and have the same properties? E.g. satellite images vs. sampling from sensors on ocean-bound vessels.
    - What kind of decision support systems or automations could the data and ecosystem support (irrespective of cost)? Which are feasible?
    - How accurate, reliable, fair, scalable would such a solution be, relative to what is done today?

Assignment

- Follow the same process as in class for an SDG area of choice. Prepare 5-6 slides and a written report covering all questions in sufficient detail and with all supporting references.

Lecture 2: AI + SDG Presentations and Discussion

- Randomly select 4-5 students to present their assignment for 5-10 minutes each, with instructor-facilitated discussion.

Assignment

- No additional assignment this week.

Topic 6: Course Project Introduction

Lecture 1: Pitching External Projects and Team Assignment

- External partners with SDG-related problems present for 5 minutes each, ensuring they cover the following questions:
    - Who are we?
    - What core issue are we trying to solve?
    - Which SDG does this relate to, and how will it be measured?
    - What is the current process?
    - What data, if any, are already available?
- *Note that we provide resources to help connect you to potential partners, and to help them to articulate their SDG-related needs in a way that fits with the class requirements.
- If too few external partners are available, or as a supplement, use the AI for Good library of open problems and previously tackled problems to give students ideas.
- Assign class into project groups of 4-6 students. We encourage you to assign randomly, with attention to distribute high-achieving students throughout the groups. However, we understand there may be students with overlapping interests where it might make sense to allow them to form groups based on these.

Assignment

- Each group is responsible for identifying 3 potential questions, and completing an Initial Research Brief for each.
    - Can be extensions of previously completed assignments.
    - Can be based on partners who presented in class.
    - Can be based on AI for Good Foundation project library.
    - Or an entirely new set of ideas.
    - Develop 2-3 slides on each potential idea, and present the one that your group believes is the most likely to be feasible, impactful, and interesting.

Lecture: Presentations of Initial Research Briefs

- 5-6 groups are chosen randomly to present their initial research brief. Students who have not previously been chosen to present in class should present on behalf of their group, to the extent feasible. Each presentation is followed by instructor-facilitated debate about the relative strength/impact of the idea.
- This week, instructors and teaching assistants meet with each group to finalise the choice of project and assign either domain mentors or scientific mentors to each team based on the topic and availability. Any team with only a domain mentor must be also paired with a teaching assistant or equivalent scientific point of contact for weekly check-ins.

Assignment

- Each group is responsible for meeting with their mentors for at least 1 hour per week, and developing an Initial Exploration Document (using the provided template, or similar) to be presented in week 8. The Initial Exploration Document should be submitted for feedback to mentors/teaching assistants with sufficient time for integration of feedback before presentations. This is a significant assignment, and foundational to the success of the projects. Each student should spend 10-15 hours working on these tasks.

Topic 7: Ethical Practices for Data

Lecture 1: Ethical Practices with Artificial Intelligence

- We encourage this class to be discussion-based. Select two case studies around the application of an Artificial Intelligence solution that had undesirable spill-over effects that systematically affected a protected group more than the average population. A good starting point for a list of pop-culture examples of ethical issues with AI interventions can be found in the book "Weapons of Math Destruction", by Cathy O'Neill. We do not recommend the book as required reading for the course due to it being written as a 'take it or leave it' opinion piece, rather than an analytical work, but it is a helpful starting point. For example:
    - Amazon's by-now famous data driven resume screening tool.
    - The COMPAS recidivism algorithm, as studied independently by ProPublica.
- Questions to explore:
    - Background—what is the problem that was being solved, and what was being done before this tool was introduced?
    - Bias—what was the bias that was eventually found, and how was the bias determined?
    - What data were available for developing the solution? How well do they match the requirements?
    - What modelling approaches were used? Any relevant implementation details that might have exacerbated the problem aside from the modelling?
    - How should such an algorithm be continuously audited for undesirable outcomes? Do there need to be changes in its deployment to allow correct evaluation?
    - How should such algorithms have been communicated to stakeholders, and what rights should stakeholders have?

Assignment

- In September of 2020 the US Federal Government released a first draft "US Data Ethics Framework"[1], whose purpose was to guide federal agencies in the appropriate use of data while promoting the public good. Read the (brief) document and answer the following:
    - To what extent does the "ethical data framework" require behaviours that are different to how we would hope US federal employees behave with respect to non-data work?
    - Are there topics missing from this framework in order to encourage inclusivity, fairness, and trust of the government's use of data on individual subjects?
    - Is it an accurate for the document to distinguish aggregate data as being low-risk to individuals, relative to individual-level data? [Consider that the majority of policy decisions are at an aggregate level, but always have individual repercussions.]
    - Develop 2-3 slides and come ready to present your finding to the class.

Lecture 2: Ethical Frameworks in Other Disciplines

- This class will look at frameworks for the use of data in settings where there are societal ramifications. We look at examples from other disciplines as a way to understand that this is not the first time humans have considered the potential for misuse of technology or procedure.
- Select 2-3 students at random to present their assignment, followed by instructor-facilitated discussion.
- Cover the different promises that doctors, engineers, and lawyers make to society when entering their profession.
    - Civil Engineers in the United States first developed a code of ethics through the ASCE in 1914.
    - In the UK, the Royal Academy of Engineering proposed ethical standards across all sub-disciplines in 2005.
    - Canada's model code for members of the legal profession, which has undergone several transformations since 1920. Chapter 7 is particularly relevant.

[1] data-ethics-framework-action-14-draft-2020-sep-2.pdf

    - Indian Medical Council (Professional conduct, Etiquette and Ethics) Regulations, 2002
- Facilitate a group activity, splitting the class into groups of 6-8 students, to explore the following questions for about 10 minutes:
    - What are the responsibilities of an AI researcher/engineer towards their discipline, their data subjects, their employer, and society in general?
- Select 2-3 groups to present their thoughts to the class, with instructor-facilitated discussion.
- NB: depending on your university's ethics resources, you may choose to (a) ask an ethics counsellor to facilitate class discussion, or (b) ask an ethics counsellor to be present in order to help students feel comfortable discussing these topics openly, without fear of retribution from peers.

Assignment

- Use the resources presented in class to consider what a skeleton AI Ethics Framework might look like for researchers and policymakers working at the intersection of Artificial Intelligence and the Sustainable Development Goals. Create a 1-page manifesto.

Topic 8: Presentations of Initial Exploration Documents

Lecture 1: Presentations of Initial Exploration Documents

- We recommend that each group be given 5 minutes to present their initial exploration document research, as reviewed by their assigned mentor, followed by 5 minutes of instructor-facilitated discussion. This should allow the class to cover 5-6 groups, selected randomly. Any individuals who have not yet presented their work to the class should present on behalf of their group. The purpose of this is to challenge groups on the impact of their proposed project, how it might be evaluated, and how they think it might be deployed. Discussing these questions openly at this stage is beneficial for the development of strong projects that students remain engaged in over the full 1 or 2 semesters.

Assignment

- From this point in the course, one lecture each week is student-led in a flipped classroom style. One group is selected for each week, and is responsible for selecting from the additional topics, and preparing content for a 30-minute lecture segment. Depending on the number of groups you have, you may want to do 2 groups per week (one in each lecture), and reduce their stage time to 20 minutes each.
- Each week, 2-3 groups should present on the progress of their project, for 10 minutes each, with questions and instructor feedback. The aim is for each group to provide project presentations 2-3 times during the course, and for every group to lead one lecture.
- NB: The original study groups needn't continue meeting from this point onwards. However, you may still find them helpful for enforcing readings and mini-assignments beyond the core capstone project requirements.

Lecture 2: Presentations of Initial Exploration Documents

- This lecture is the same as the previous lecture, allowing several groups to present their initial exploration document to the class.

Assignment

- Groups should be meeting at least 2x per week to coordinate work on their project, as well as at least 1x per week with their assigned mentor. It is expected that each individual should be spending on the order of 6-10 hours of non-class-time each week on the group projects. Readings and mini-assignments related to class content may still be required, but will be minimal.

Additional Themes

Each theme is planned to span two weeks of class, with two flipped classroom exercises tied to each, and allowing enough time for other activities that need to take place relating to projects. Thus, the material presented here is meant to cover the residual 2 hours of class time, much like the core 8-week schedule presented above. As a result, you may also choose to switch around any of the following with the core modules.

A. Evaluating Machine Learning Models Effectively [4 hours]

- This module aims to provide students with a basic framework for thinking about the "performance" of Machine Learning models in practice. We cover:
    - Data selection and auditing. How do you check that your data sets are fit for the purpose that you intend? How do you ensure your data sample has the same properties as the real world deployment you imagine?
    - Traditional methods of training, test, and evaluation sample selection, cross-validation, ROC-curves, F1 and other scoring criteria, etc.
    - What is an error? Thinking about fundamental concepts of risk, cost-benefit analysis, in the context of training and using Machine Learning models.
    - Error bars and statistical significance—to what extent can we believe that the evaluation metrics we observe actually represent the likely operating characteristics of our model in the real world? How do we calculate this certainty using statistical tools?
    - Continuous Evaluation and Dynamically Updating Models. Real world models are deployed in an environment that is constantly changing. The cultural context changes, other policy interventions happen, data collection methods change/improve. Our models

need to be regularly re-evaluated and updated, if not continuously. However, the pressure to counter environmental dynamics needs to be balanced with a conscious evaluation procedure to avoid unintentionally introducing performance issues later on.

- How good is good enough? The Machine Learning literature encourages publications that show incremental improvements in well-known domains, but AI for the SDGs requires 'good enough' performance on problems that have no existing model-driven solutions. What is good enough relative to existing procedures, societal expectations, and the inherent viability of the project?
- Interpretability trade-offs. What is it worth to have more explainable models? How 'explainable' do they need to be? What are the common misconceptions about "explainable models"? [For example, the ability to assign a weight to a feature in a model's decision-making process does not mean that the feature is independent].

B. Thinking about Privacy and Security in AI [2 hours]

- This module considers the idea that models might 'leak' personally identifiable information, even when steps have been taken to de-identify data sets. What can be done to ensure that models are representative, without exposing the underlying data subjects in your training data to re-identification and other risks?
- We primarily look at issues related to "Privacy Aware Learning" as proposed by Michael Jordan in 2014.[2]

C. What is the Economic Impact of Artificial Intelligence? [2 hours]

- This module considers how to understand the effect of the development of Artificial Intelligence on the economy, in terms of productivity, scalability, employment, measures of inequality and income distribution, and international trade. This is an under-researched area that is gaining more and more interest from

[2] https://dl.acm.org/doi/pdf/10.1145/2666468

academics and policy makers. This has important ramifications for the Sustainable Development Goals directly and indirectly.
- We look at these issues partially through the lens of recent academic work in this area.[3]

D. AI and Policy in Practice: Designing Policy Interventions

- This module explores how governments and NGO's go about the policy design and implementation process, and key issues with the implementation of AI-supported interventions. To what extent is AI critical to the design of a policy implementation? We consider elements such as stakeholder processes, auditing of technology infrastructure and security, how/where should infrastructure be deployed, how will it be maintained? How can policy interventions be appropriately evaluated relative to non-AI interventions? And a variety of their related themes as well as real-world case studies.

E. National AI Strategies

- More and more countries are developing 'National AI Strategies', with more than 50 such national-level documents tracked by the OECD Policy Observatory as of 2021. This module will take a series of case studies of these strategic initiatives, and look at them from a policy perspective. Only a few of these national strategies make direct policy recommendations, with many of them boiling down to 'invest in AI', and 'invest in increasing the number of AI skilled workers'. We leverage the AI for Good Foundation's AI Policy categorisation to understand the different national priorities and how they fit with these documents, and take a deeper dive into how a national AI strategy might interface with the Sustainable Development Agenda and Goals.

F. Useful ML Tools and Resources

- This hands-on module introduces a series of data and ML research support tools, libraries, and frameworks that can help accelerate development, collaboration, and communication during the group

[3] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3651052

projects. The tools will be presented interactively, with small group activities to leverage each for a particular objective.

G. Perceptions of AI in Society

- This module explores the communication and perceptions of AI in society, from news media, to governments, prominent individuals, companies, scientists, and movies. How do people perceive the future of work? The future of health? The future of education? What are the main fears and apprehensions? What are the positive expectations that people have? To what extent is the general population receptive to and educated about Artificial Intelligence, and how does this interplay with policy and government operations? Finally, how should we communicate issues and opportunities around Artificial Intelligence in order to build an informed and inclusive stakeholder public?
- This module leverages resources connected to the AI for Good Foundation's Global Perceptions of AI Survey.

Course Planning

After weeks 8-10 of the core course themes (shown above), the course can be crafted in a fully modular way. Any of the additional themes below can be incorporated, in whatever order makes the most sense for your students. You may even find that you want to include these topics earlier in the course, before week 8. All is up to you, and we will soon be offering tools to make the modular creation of tailored course plans super simple. Each week you will continue hosting group presentations and flipped classroom experiences.

For shorter (1 semester) courses, you may want to do the following:

- Core 8-week schedule as shown.
- Start group assignment and project planning from week 2-3.
- Reduce number of individual assignment presentations, to allow time for project presentations.
- Make the course 50% individual grade, 50% group grade.
- Introduce the Initial Research Exploration document presentations as a “midterm” with invited external stakeholders from other departments.
- Drop the flipped classroom approach, since too few groups will have an opportunity to present.
- From week 6 to the end of the semester, organise project deliverables roughly as follows:
    - Core data collection and system design (storage, databases, infrastructure components, UX considerations, deployment).
    - Machine Learning development plan.
    - Modelling, Infrastructure Development.
    - Proof of concept deployment plan (recruiting a test population).
    - Evaluation and Project Report.
    - Final Presentations to stakeholders, and submission of project packet with all resources

- If you have time for additional lecture material, prioritise the continuous evaluation and interpretability of models, AI and Policy in Practice, and the Public Perceptions of AI themes.

The included Core Syllabus follows the 1-Semester Course Planning. Because of this, you will notice that weeks and topics do not necessarily line up in the syllabus.

For longer (2 semester) courses, you have more flexibility. You may consider the following:

- Schedule as shown.
- At the end of the first semester, invite external stakeholders from relevant departments and the community (including remote guests) to attend presentations of each group's project progress. Think of this as a 'pitch-week' to get more comprehensive feedback, and more exposure to communicating their ideas outside of a student setting.
- Consider spending more time on some of the core themes:
    - Ethical Practices, Machine Learning Evaluation, Policy Evaluation, Causality and Experimental Design, the Sustainable Development Goals and Data.
- Consider more time for individual assignment presentations (i.e. more students present each week) and flipped classroom exercises, group exercises during class time, etc. You could run the 8 week schedule over 12 weeks by spanning 3 classes per theme instead of 2.
- From presentation of the Initial Exploration Documents, until the final project presentations, time on project development should be spent roughly as follows:
    - Core data collection and system design (storage, databases, infrastructure components, UX considerations, deployment).
    - Data Evaluation and Auditing.
    - Machine Learning development plan.
    - Ethical Audit of ML plan
    - Model Evaluation Plan Pre-registration. Defining success and expectations.
    - Modelling, Infrastructure Development.
    - Proof of concept deployment plan (recruiting a test population).

    - Evaluation and Project Report.
    - Final Presentations to stakeholders, and submission of project packet with all resources.
- Consider the creation of an ‘accelerator programme’, funded by the university, to allow the top-n projects to continue development with additional mentors through the summer after the second semester. Talk to the AI for Good Foundation about supporting the creation of such a programme. Groups could be evaluated by the AI for Good Foundation Council, or by a local set of stakeholders who attend the final presentations.

Contact

For questions and additional resources, please reach out to info@ai4good.org.

About the AI for Good Foundation

AI for Good is driving forward technological solutions that measure and advance the UN's Sustainable Development Goals. We create impact by bringing together a broad network of interdisciplinary researchers, nonprofits, governments, and corporate actors to identify, prototype and scale solutions that engender positive change. Founded in 2015 by a team of Machine Learning and Social Science Researchers in the US and Europe, AI for Good is headquartered in Berkeley, California with an international network of core team members, partners and volunteers supporting our work.