Introduction

* Draft deadline is midday Friday, with no extensions possible, so make sure you get it done asap.
* Late drafts might still get a review, depending on how quick you are compared to reviewers.
* Tutorial 3 is this week, remember to submit your notes ahead of time.
* This is the start of the final section of the course.
* First, a brief recap.
* The first portion of the course covered some motivation for why it is important for all of us to be concerned about ethics in computer science, linking in to wider concepts like the impacts of distribution of Power and how Responsibility falls differently on different people.
* The middle portion of the course looked at some specific ways in which technology can cause harm, whether intentionally or unintentionally, such as the misuse of their Data, the propagation or introduction of Biases, or overlooking the value of real Humans in a system.
* This final section will be about some possible mitigation techniques, and Meenakshi’s lecture served as a nice bridge, highlighting the challenges real engineers face when trying to balance between their jobs and potential harms
* Most of the lectures in this section will be things that are within your power to enact, even if that requires some coordination with other people. We will talk about Design Frameworks teams can implement, Personal Attributes we can cultivate in ourselves that might allow us to approach our work more ethically, and the value of workplace organising.
* This lecture is going to cover some of the ways beyond specific individuals or teams that efforts can be made to encourage ethical tech development.
* I’ve called this section Leadership as a catch-all term, because these are generally the sorts of solutions people don’t put in place for themselves, and are instead recommended or implemented by people with authority or power.
* The power to enact these things can be collective too, so even without being in an explicit position of leadership they are still things we can push for, as in the example of the Amazon Employees for Climate Justice campaign.
* So let’s have a look at some of the different forms of leadership in vague order from least to most formal.

Ethics Guidelines

* To start with, there are ethics guidelines, which is a loosely defined category basically consisting of any set of advice on how to work or what technical systems should or shouldn’t do.
* A lot of companies will have some kind of guidance internally at least, included in their training materials, stating the kinds of high-level opinions the company culture holds about its work.
* These kinds of guidelines are often very vague, and they are not meant to be rigid rules but more advice for people to follow. A perhaps extreme example of this might be Google’s “don’t be evil” motto. Maybe indicating the right overall direction for the company’s work, but certainly not intended to be a testable standard against which everything Google does will be measured.
* The benefit of good guidelines is in some way in their vagueness, though. They are less focused on ruling about every possible case and more about the attitudes or approaches that should be taken to the work. A well-made set of guidelines should be teaching the people applying them how to behave more generally (perhaps conveying desired Personal Attributes or Design methodologies), and leaving the specifics of implementation in individual cases to them.
* Notably the “don’t be evil” statement was in what Google called a code of conduct, exemplifying that the boundaries here are not fixed. Also notably, they removed it in 2018. Probably for different reasons.

Codes

* Codes of conduct, such as the ACM Code of Ethics, might be thought of as more specific formalised versions of guidelines.
* They focus on the behaviour expected from people within a discipline, and they enumerate this into a set of concise, distinct expectations. This has the advantage of making it easier to contrast your own or another person’s actions with the expectations set out in the code, and to know which points are or are not being met.
* On the other hand, the concision needed for a snappy code of conduct means it is unlikely to contain additional instructions for particular processes, or justifications which explain to the reader *why* they should care about following each section of the code.
* Guidelines and codes of conduct are both also limited in terms of enforcement. The focus is on being easily understood and followed, rather than setting out a series of tests, which means an external observer can often see only that there is some kind of guideline in place and much less how much employees have taken it to heart or how strictly it is followed.
* Probably worth noting there are two related kinds of enforcement we’re interested in here. One is how easily individuals or their actions can be compared against the code of conduct and the other is how easily a similar thing can be done at company level.
* One of the risks of relying too heavily on guidelines and codes of conduct is that it is pretty easy for a company to *state* that they have one, but interrogating whether they actually make any concerted effort to follow it is another question entirely.

Standards

* Standards (which many of us covered in SEPP) come from a different direction. While codes of conduct or guidelines are usually picked up voluntarily by individual organisations, Standards are proposed and agreed upon often at the level of a whole industry.
* They are generally much more specific (and therefore also testable) best practice rules for how to do something. Sometimes this is rules about the requirements for a product that must be produced to a specific quality, such as a smoke alarm or cricket ball, and sometimes it is rules for whole processes, such as Quality Management.
* In many areas, Standards often come from the industry itself, because even competing companies recognise the need for certain things to be reliably done in the same way. Unicode, the platform-agnostic way text (and just as importantly emoji) are encoded across devices, is a standard. The methods by which aircraft engines are tested for safety are standards. Anywhere people see the need for a products or systems of a certain type to all be judged the same way and guarantee certain capabilities, Standards are used.
* Standards for a lot of artificial intelligence are still relatively few and far between, partly I suspect because it takes a long time for companies to recognise the need and agree on what is good for them. Standards for ethical AI, or for ethical tech development practices, are still rarer, but it is certainly one potential future avenue.
* Explainability is one area where this has been proposed to be useful. Creating a Standard for what information should be retrospectively retrievable from an AI decision making system, and how, would put a hopefully well-defined onus on developers to not produce any system that it was impossible to interrogate should unexpected behaviour occur.
* Standards also have their down-sides of course. As noted, because they are generally industry-driven, they can take a long time to be agreed upon and finalised. They also obviously will generally fail to cover anything where a significant portion of the industry fails to see a business case for being restricted, unless some other outside pressure like campaigners or government intervene.
* They also by their nature as being more specific often find it challenging to cover all potential cases. This might lead either to Standards only for very constrained tasks, like what information a self driving car must be able to report post-crash, but not for more open ended problems, like how to behave in a situation where a crash is likely.
* As for how obligatory Standards are, it varies. Some Standards will be essentially optional, but with heavy industry incentives that motivate people to follow them. Think of champagne, which is only allowed to be called that if produced a particular way. You can produce something very like champagne by not following the rules entirely, but you won’t be able to label it the same. Once a company does agree to a Standard, though, they will be more tightly held to its rules than say, if they announce they are following particular guidelines or code.

Legislation

* The law and Standards have some reasonable overlap, in the sense that Standards might be designed to guarantee any system following them is also meeting its legal requirements, or a law might specifically refer to a standard as being a requirement.
* Law is, probably unsurprisingly, the most formalised version of the systems I’m grouping together here as Leadership. Within the region in which they apply (which for international law is essentially everywhere), laws are a relatively unavoidable set of constraints on goods or processes.
* The law applies to everyone in a region, whether they choose it or not, and this obviously has the advantage that law can theoretically impose rules on industries whether or not those industries would choose to be restricted in that way, although in practice an expensive legal team might still make some powerful players harder to pin down.
* The requirement to be enforceable, however, is the reason that the law has to be specific and precise in its wording, which again makes it hard for it to cover more open ended issues, at least without relying heavily on interpretation by courts. It also makes the law incredibly inaccessible to the casual reader, which means that contrasted with the other kinds of leadership covered so far it is likely to do a very poor job of explaining what it actually expects of people, let alone why that is important.
* This means a down side of law can be that it often means a reliance on the specific expertise of individuals trained in legal matters, who might not even be at the company. This can lead to the rules being seen as just a series of unexplained “allowed”/”not allowed” decisions, which is especially problematic for complex ethical issues where deeper understanding and involvement of everyone in the process are needed to cover all of the nuance.

Ethicists

* Finally, I want to talk about a different kind of trained individual a company can employ in a kind of leadership role: the ethicist.
* It is becoming more common, especially at larger companies which both have the budget for it and the need to improve their public image, to hire individual or even teams of ethicists.
* The exact role of these can vary from place to place, but often their remit is to have oversight over all of the work the company is doing and to provide analysis and advice for what to do differently.
* This has the advantage of being flexible, because the Human ethicist doesn’t have to be designed for predefined situations like Standards or Laws, and they hopefully can also do a good job of explaining or teaching important concepts to other people in the company.
* The effectiveness of this role is incredibly dependent on the specifics of its implementation, however. One mistake that it seems is often made is giving the ethics team no real power or influence, to the extent that whatever they advise can just be ignored by the rest of the company. Similarly, if they are just a separate team without proper integration into other employees’ work or processes, then they will come to be seen like the Legal Advisor: just a yes/no machine sometimes getting in the way of work, or occasionally giving company-wide presentations and otherwise being forgotten.
* Unless a company employs a huge number of ethicists, overseeing every technical decision made, it will still be important for everyone else working at a company to have an understanding of, and appreciation for, the ethical dimensions of their work. Ethicists can then act as domain experts, much like the company might have a specific computer vision, or databases team, who can work collaboratively with everyone else on particularly challenging questions or problems.

Conclusion

* Of course, none of these practices is likely a solution on its own. Hopefully by appropriate application of each, they can cover for each other’s weaknesses.
* Remember when you’re interacting with them to treat them as a part of that wider ecosystem.
* Maybe a company code of conduct wants something from you it doesn’t give instructions for, and a Standard might help.
* Maybe a law seems unnecessarily restrictive, but an ethicist might be able to explain the reasoning.
* Ultimately, however much leadership is shown by people in power across the field, in the end it is the people doing the development or deployment work who have to be mindful of their actions and responsive to new situations. They just don’t have to, and indeed shouldn’t, try to solve every problem on their own.