Rohan Anand
2/14/23
DS 380
Professor Wildman

A Deontological-Based Code of Ethics for a Machine Learning Engineer

In the age of revolutionary, fast-paced technology such as AI, the rate these programs are integrated into our lives outpaces the ability the government to protect citizens from negative consequences. From the false arrest of a black man due to facial recognition to the death of a woman caused by a self-driving car, officials are unable to prevent the harm the technology is capable of perpetrating. As a result, the ethical burden of developing AI that does not harm its users fall upon the engineers. This is the career I will pursue. As an aspiring machine learning engineer, I will be responsible for vetting data, designing the model, and developing metrics to measure accuracy. It is my obligation to ensure that these safe-critical systems serve clients equally, as I "have the potential to adversely affect an 'increasingly large and diverse clientele by failing to act responsibly, fairly, timely and appropriately" (Tavani, 90). To achieve this, I will adhere to a code of ethics, based on deontology, by formulating potential ethical dilemmas and proposing principles to resolve them, and in totality, constructing a philosophy by ethically approaching of each part of the modeling pipeline.

My ethical philosophy follows deontology, which is founded upon the principle that obligation to others is the foundation of morality. It centers around individualism, claiming that individuals have moral worth, and, should not be used as a means to an end, unlike its contemporary, utilitarianism, which is aims to maximize the happiness of the majority by sacrificing the minority. Each individual has a duty to act ethically to others, regardless of the consequences. The central belief is to treat others the way you would like to be treated. This system supports my career, as being the one with intimate knowledge of AI models, I carry the responsibility to inform others about their limitations, even if it sacrifices profits for employees/shareholders to protect users. Also, I have a duty to ensure these models are able to produce accurate results for every group, as if a model were used on me of my life, I would not want to be discriminated against.

To achieve this, I will

1. **Use data with informed consent**

   The rule for AI models is that more data is better, as it helps uncover more patterns, increasing its accuracy. So, companies collect lots of data, enhancing their services, such as retention rate for advertisements. However, this desire for improvement can exceed ethical bounds. For example, social media companies collect invasive data on individuals, such as location data, device information, etc. It can be used with malicious intent, similar to TikTok employees using IP data to find employees leaking secrets, demonstrating that companies can use this data for their own purposes. If I was asked to use invasive data to construct a model, I would object for two reasons: as one who has access to the data, I am obligated to report data that users are not aware is being collected, as they deserve to consent to give it because they own their own data, and I would not want my data exposed by others without a chance to object.

2. **Validate data for fairness/equality for all**

   As AI becomes integrated with fields such as policing and diagnostics, more demographical data is being collected. However, these models cannot verify if the data is equitable, and can produce false predictions due to its inequality. For example, facial recognition or diabetes prediction can fail on minorities leading to false imprisonment or unnecessary health scares due to sample size, proportion of represented groups, etc. To ensure accurate results for all, I will make sure the data comes from a representative sample, considering all races, socioeconomic statuses, etc. as it is my duty to inform users about the possible limitations of models to prevent harm. As a minority myself, I would also like to be wary if the model is not accurate for people like me.

3. **Build transparent models with clear reasoning, in cases of human usage**

   Because AI cannot reason, it can produce conclusions based on false premises, such as racism. This is problem with deep learning, which uses a neural network. The way neural networks adjust its weights, the millions/billions of parameters, the non-linear activation functions, etc. means its result is uninterpretable by humans. So, using its results can have unintended consequences. For example, arresting incorrect suspects, which results in psychological trauma and legal troubles. Although utilitarians argue that a facial recognition system captures more criminals than civilians, I would not build an AI model to do this because it undermines an individual's privacy and allows the government to circumvent collecting evidence/investigating. Similarly, I would want a model that is used to accuse me to have explainable results.

4. **Develop AI models that computationally efficient**

   Compared to traditional machine learning, such as decision trees, more advanced AI models that utilize deep learning take longer amounts of time to train, resulting in greater energy expenditure, and therefore a greater carbon footprint. The carbon emissions produced by training GPT-3 is equivalent to what 13 American's produce in each year. This adds up, across numerous models, across numerous companies. These emissions can damage the Earth, harming the environment that other individuals like myself rely upon for survival. As a result, I have an obligation to them to reduce emissions to preserve the environment as much as possible. I am obligated to rely on more computationally efficient, energy efficient methods of modeling as I would also like others to preserve the environment.

5. **Produce accuracy metrics that reflect the performance of the whole**

   As mentioned above, some algorithms have better accuracies based on certain groups of people due to the demographical data. This occurs when diagnosing skin cancer, where there exists more data for Caucasians, even though non-Caucasians can receive it. So, an algorithm used to detect skin cancer, will have a higher accuracy with those with less melanin. The accuracy can be skewed in their favor through weights based on sample sizes, neglecting results for minorities. However, I have an obligation to report the accuracy separately for all groups instead of averaging, noting that it is due to the lack of data. I also would not want an algorithm that can miss diagnoses used on me.

These 5 principles are the pillars of a deontological-based philosophy that will allow me to produce AI models that are ethically once I become a machine learning engineer because of the consideration for obligation for the its users at each step in the data pipeline.

Works Cited

Tavani, Herman. *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*. 5th ed.