IFPMA | Ethoscope

AI Ethics in practice: Use cases

Use case 01

A pharmaceutical company develops software that uses Artificial Intelligence (AI) to help physicians with the treatment of patients with cardiovascular diseases (coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions).

This software is available in several hospitals around the world, and it makes treatment recommendations based on historical clinical data from patients coming to the hospitals (previous treatments received, past conditions, basic personal data - age, gender, nationality, blood group, etc., historical expenditure on healthcare - i.e., pays for a private insurance yes / no). In other words, this AI solution is used to directly determine patient treatment options.

The software uses the above-mentioned data to recommend treatment options to physicians treating patients. Physicians find it is useful in their daily practice and therefore, the software helps to increase patient's benefit and improves overall healthcare.

Following the company's Artificial Intelligence policy, researchers analyse the software recommendations some months after launching it and find out that the software was making cheaper treatment recommendations to People Of Colour (POC) patients (African or afro-descendants individuals), compared to those made to Caucasian patients. This finding was brought up to the company's AI Ethics Committee, to the attention of senior management.

To mitigate this race bias, researchers continued to analyse the algorithm, and found out that the AI model was wrongly identifying POC patients as having milder cardiovascular diseases, therefore making cheaper treatment recommendations than those made for Caucasian patients.

To understand the origin of this outcome in the model, researchers in the developing company analysed the algorithm and the datasets that fed and trained it. They realized that the algorithm was ignoring the variable "annual income", and only considering "historical healthcare expenditure" to determine treatment recommendations. Therefore, the algorithm was omitting the fact that POC patients have an average lower annual income than Caucasian patients, which is why the "historical healthcare expenditure" for POC patients is lower than Caucasian patients.

The company researchers corrected the algorithm training by including "annual income" as a core variable to train the algorithm, documented this finding and the remediation actions in the software records. The company communicated the situation, remediation steps and future plans to the AI Corporate Ethics Committee to ensure the software was beneficial to all patients and helped increase healthcare overall.

Which AI Ethics principle is addressed in this use case?

Fairness and minimisation of bias

In this case, the risk of bias was not evaluated correctly at the initial stage. The AI solution behavior showed it was trained in a way that discriminated POC patients.

The monitoring of datasets used in training and control measures enabled the identification of the bias and allowed corrections to be put in place.

Use case 02

A pharmaceutical company focused on respiratory diseases develops an app, called BreathePro, for patients with Chronic Obstructive Pulmonary Disease (COPD).

BreathePro is intended to help patients manage their condition, offering tips and recommendations based on the information entered by the patient (breathing quality, tiredness, sleep quality, level of anxiety, overall emotional state, etc.) With all of this information, BreathePro recommends guided meditations, breathwork exercises and other general lifestyle recommendations to improve the patient's quality of life.

One of the outcomes BreathePro delivers is the possibility of suffering lung cancer. This outcome only appears if patients input certain symptoms like sustained high levels of breathing difficulty, anxiety and tiredness over the past month. In such cases, BreathePro also recommends the patient to schedule a visit with a physician.

A specific COPD Patient's Association receives a complaint from one of its members, saying that her anxiety levels increased after receiving a message from BreathePro claiming that there was a possibility of her suffering from lung cancer. She claims that her COPD symptoms worsened with the anxiety she felt after this message, and that when she visited the doctor to confirm it, she was just diagnosed with a severe anxiety crisis. No lung cancer risk was identified by the physician.

The COPD Patient Association gets in contact with the owner pharmaceutical company to report this incident. Following the key accountability principle, the company first got in contact with the physician and the patient who suffered this situation to apologize and thank them for reporting it. The company then had data scientists analyze the BreathePro algorithm to remove the "possibility of suffering lung cancer" from its chatbot, as it raised unnecessary concerns for patients using it. Instead, data scientists modified the chatbot to add a more cautious message that indicated patients to visit their doctor if certain symptoms were inputted by the patient.

Which AI Ethics principle is addressed in this use case?

Accountability

In this example, the pharmaceutical company that developed BreathePro was accountable for the impact of the app on the distressed patient. Therefore, they responded to the physician and patient's concerns, and analysed the algorithm that built the AI solution to understand its behaviour and identify the ethical issue.

Explainability (in Transparency, Explainability and Ethical Use)

BreathePro send the lung cancer message to the patient because the entered data matched the requirements for this message to be delivered. Data scientists analysed the algorithm to explain its behaviour, correct it and avoid this problem happening again.

Use case 03

A pharmaceutical company focused on neurological diseases has partnered with a neurotechnology startup to develop brain microchips connected to electrodes to treat Parkinson disease.

Both companies start the BrainTech project, to develop a microchip to be implanted in the patient's skull, with electrodes expanding to the brain region that is damaged due to Parkinson's disease. When the device is ready and has tested positively in lab trials, the BrainTech project makes the next step to clinical trials with humans. The device is tested in real Parkinson patients to see how effective it is in reducing disease symptoms and progression.

After years of running the trial, one of the participating Parkinson patients reports experiencing depression symptoms. He claims he has been feeling depressed for the past two years and that he is starting to have suicidal thoughts.

The patient also claims that his sense of self has somewhat been altered after several years of wearing the device and that he has begun to wonder whether his depression was caused by the device itself, if it was a logical consequence of suffering Parkinson for so long or if it reflected something deeper about himself.

The companies owning BrainTech (pharmaceutical and start-up) document this important report and follow up with the BrainTech Ethics Committee. After deciding there is a high risk of patient harm and that it possibly correlates with the device implanted in his brain, both companies decide to remove it from the patient. The device is removed from the patient's brain, he is then put in intensive psychiatric care and depression symptoms are regularly monitored.

BrainTech Ethics Committee followed their core principles of human control and accountability and decided that individual integrity prevailed over scientific progress. They documented the incidence and the remediation measures adopted to prevent such incidents repeating in the future. Among others, data scientists analysed the datasets that fed the microchip to work with neurologists to try and identify the root cause of this serious incidence.

They also documented the important fact that they were unable to demonstrate if there was a causal relationship between the AI model in the device and the depression symptoms, to follow their also core principle of transparency in front of unexplainable AI. After several analysis, owning companies decided to stop using the device on humans.

Which AI Ethics principle is addressed in this use case?

Empowering Humans

BrainTech was initially designed to help Parkinson's treatment and with the hope to increase patients quality of life. The unexpected outcome on the affected patient made the owning companies realize the AI system somehow took over human control and autonomy. This is what made them take the decision of not using such device in humans. This is how they respected the principle of AI respecting human rights and dignity.

Explainability (in Transparency, Explainability and Ethical Use)

It was not possible to explain what made the affected patient experience depression symptoms and suicidal thoughts. However, as the patient experienced this significant change in his behaviour after having the device implanted, it is logical to suspect a possible link between the two elements.

A useful work around to manage unexplainable AI is to have a system in place from the very early start of the AI design that helps developers predict behaviours that could led to ethics dilemmas.

In addition, pharmaceutical companies working with unexplainable AI should put an extra focus on transparency, human control and elimination of bias.

Use case 04

A pharmaceutical company starts a project with hospitals around the world to gather genetic data from patients diagnosed with colorectal cancer to identify genetic markers, in the hope of understanding and research the genetic base of the disease.

Both the company and hospitals get the anonymized genetic data from patients, and share the aggregate data with a gene bank so that they analyse it in search of patterns and possible colorectal cancer genetic markers.

After the project is ongoing for months, it is disclosed that a group of employees from the gene bank, the hospitals and the pharmaceutical companies had been sharing this data to health insurance companies. These companies had worked with computational scientists to disaggregate this data and offering different offers or conditions to their clients based on their genetic information.

As this misuse of the AI system was unacceptable, an official investigation was conducted by an external audit company in collaboration with the Department of Justice DOJ). Senior management from the companies involved

(pharmaceutical company, hospitals, and gene bank) were held accountable for the incident and gave a public explanation of the events that had led to such unethical use of AI systems.

The project was terminated, and all three companies worked with the external audit firm to establish robust internal processes that established strong controls in place and prevented this situation from repeating again. This included writing an AI Ethics policy, establishing an official AI Ethics Committee, actively training employees on AI Ethics and launching a whistle-blower system, among others.

Which AI Ethics principle is addressed in this use case?

Transparency and Ethical Use (in Transparency, Explainability and Ethical Use)

The way patients data was deliberately disaggregated and shared with health insurance companies is clearly a non-transparent, unethical way of using AI systems.

Privacy, Security and Safety by Design

The way personal data from patients was deanonymized and shared with another company completely disregards applicable data protection regulations and the basic right to consent.

Accountability

It was when the situation came to light that the involved companies were held accountable for what they had done and acted accordingly, working with the competent authorities to put the necessary measures in place to prevent such incident repeating again in the future.

Use case 05

The HR department of a pharmaceutical company launches an AI based program that scans thousands of resumes from job applicants to make talent search more efficient.

The AI based program, called iHire, reviewed all the resumes received to every job offer from the pharmaceutical company and used the embedded algorithm to rate resumes from one to five stars. Only resumes with one and two stars were delivered to HR specialists to continue the selection process; resumes rated below two stars were discarded right away by the algorithm.

After some time in the market, data scientists in the pharmaceutical company analysed the algorithm behaviour as part of their regular monitoring plan and discovered that 95% of selected resumes came from male candidates.

To understand and correct this gender bias, data scientists further analysed the algorithm and discovered that it was being trained on common words appearing in resumes, penalizing resumes that included the word “women” or “female” (i.e. “women’s athletics club coach”). As these words were penalized, women candidates’ resumes were being discarded by the algorithm. As a result, nearly all the resumes that made it through the algorithm came from male candidates.

After this bias was detected, the HR team communicated it to the AI Corporate Ethics Committee. The decision was made to 1) correct the algorithm training to make it gender neutral and 2) updating the current recruiting policy, adding AI specific considerations that included a risk assessment to prevent ethics incidences like this one repeating again in the future.

Which AI Ethics principle is addressed in this use case?

Fairness and minimization of bias

The way the recruiting AI system was behaving clearly displayed a gender bias. Having an official monitoring system in place helped the owner pharmaceutical company spot the incident and quickly start looking for corrective actions.

Accountability

As the company had a governance model in place, they could use the results driven from the regular monitoring plan to identify the issue and work on corrective actions to avoid it happening again.

Use case 06

A pharmaceutical company develops an AI solution in its CRM system to automatize the tiering process of medical experts. The medical and digital team partner up to develop this solution, hoping it will make their work more efficient.

The solution reviews data from Medical Experts that have interacted with the pharmaceutical company (peer recommendations, publications, research and clinical experience, society memberships/appointments, editorial board membership), and, considering the medical experts basic personal data entered in the CRM (country of origin, healthcare speciality, age, etc.) classifies medical experts in different tier levels.

When the AI solution was developed, the designing team, together with medical and IT, worked with Compliance to build a risk assessment tool that could help predict possible misbehaviour of the solution. In addition, the medical and IT departments actively monitor the algorithm behaviour during the first two years after launch to detect any bias or incorrections to amend. In one of the monitoring exercises, the team detects that the algorithm is classifying a high percentage of MEs in countries where prescriptions are higher as Tier 1. The algorithm does not behave this way with MEs that are based in countries were prescriptions are not so high.

After communicating this bias detection to the appropriate teams within the company (commercial, medical and IT senior management), the value “country” is modified in the algorithm training model so that it is not dominant value, and the tiering process is done in an objective way, independent from prescription levels.

The team appropriately document this finding and remediation actions taken, to prevent it from happening again in the future.

Which AI Ethics principle is addressed in this use case?

Accountability

Because there was a robust governance model in place that allowed the teams to put the measures in place to predict AI behaviour, the ethical misuse of the system was identified. Having a monitoring plan that allowed the team to regularly check the system behaviour allowed them to easily spot the AI irregular behaviour.

Fairness and minimisation of bias

The above good practices in place (governance model and regular monitoring) helped the pharmaceutical company identify the bias that was being performed by the AI system. Having this internal governance and monitoring process in place helped the team quickly identify and correct the bias.