## IBM Gastspreker

IBM ofwel International Business Machines is een wereldwijde technologiebedrijf dat bekend staat door het leveren van hardware, software en diensten over de hele wereld. Het bedrijf werd opgericht in 1911. Het bedrijf is bekend om zijn leidende rol in de ontwikkeling van computers en andere technologieën, zoals kunstmatige intelligentie en Quantum computing.

### IBM Research

IBM heeft in de tijd zeer veer baanbrekende computertoepassingen ontwikkelt waaronder Deep Blue. Deep Blue is de eerste computer die de mens verslaat in een wedstrijd shaken. Dit doormiddel van game theory die machine had toen 256 processors om dit te solven. Hoe heeft deze computer eigenlijk kunnen winnen. De computer kan meerdere zetten per seconde te bekijken dan de mens. Het systeem gebruikte een dataset van eerder gespeelde games. Deze computer is doorontwikkeld naar het project Debater. Debater is een AI-systeem dat bedoeld is voor debatten te voeren tegen menselijke tegenstanders. Het systeem is ontworpen om menselijke debatteerde te helpen bij het verzamelen en beoordelen van informatie, het genereren van argumenten en het formuleren van hun standpunt. Dit systeem maakt gebruik van een NLP-systeem zowel offline als online analyses en werkt aan de hand van knowledge graph.

Nog een project van IBM is de Mayflower. De Mayflower is een volledig autonoom schip die bestuurd wordt door artificiële intelligentie, die vaart op zonne-energie en biofule. Er zitten drie verschillende AI-systemen op het schip. Het eerste systeem is de AI-captain dit heeft 6 camera’s, een radar, AIS en weather information aanboort. Bij dit systeem is er een probleem dat ze hebben moeten oplossen. Als de boot aan het varen is op de golven kan de boot niet altijd alles zien, ze kunnen achter een golf zitten waardoor het AI-systeem denkt dat ze gaan botsen maar eigenlijk niet. Het tweede systeem is de operational decision manager. Dit systeem genereerd risico maps, het waarschuwt u van onveilige situaties. Dit systeem heet de macht om de AI-Captain te overrulen en kan overgaan tot een volledige stop. Het laatste systeem is de Decision Optimization, dit zal de zoute zo optimaal mogelijk maken. Het kijkt naar de best mogelijke route, de meest zuinigste zoute, meer zon, … Voor de route te bepalen is er gebruik gemaakt van lineair programmeren.

### AI/ Non-AI

Ai Lifecycle Automation bestaat uit vier delen die in een cirkel kunnen worden uitgevoerd. Prepare, Build, Run en manage. In het volgende deel zullen we het meer hebben over Het Manage gedeelte van de Ai Lifecycle. Om dit beter te verstaan gaan we aantal begrippen en problemen die de AI kan ondervinden bespreken.

Fairness: hoe fair is iets. Dit kan je van twee standpunten bekijken. BV aan de hand van deze foto hieronder. Je kan zeggen dat de rechtse fair is voor iedereen. Maar je kan ook zeggen dat de linkse fair is omdat iedereen op de zelfde hoogte staat.



Bias: is een fout of afwijking van de waarheid in de resultaten van de studie of experiment. Bias is niet per se slecht of niet slecht in een ai model. Bij het trainen van het model komt bias het meest voor.

Drift heb je ook, dit is het AI-model gebruiken waarvoor het niet is op getraind. Dus de productiedata is niet dezelfde als de getrainde data van dat AI-model.

Correlatie verwijst naar de relatie tussen de invoerkenmerken en de uitvoervoorspellingen van een model. Een positieve correlatie betekent dat wanneer de waarde van een kenmerk toeneemt, de voorspelde waarde van de uitvoer ook toeneemt. Een negatieve correlatie betekent dat wanneer de waarde van een kenmerk toeneemt, de voorspelde waarde van de uitvoer afneemt. Correlatie kan worden gebruikt om te identificeren welke invoerkenmerken de grootste impact hebben op de voorspellingen van het model en kan worden gebruikt om de prestaties van het model te verbeteren.

IBM heeft heel wat diverse toolkits voor het leren en snel starten met AI fairness, deze zijn natuurlijk opensource.

De volgende techniek die wordt gebruikt bij AI-model is LIME (Local Interpretable Model-Agnostic Explanations). LIME is een techniek voor het verklaren van voorspellingen van machine learning-modellen. Het maakt gebruik van een benadering die agnostisch is voor het specifieke model dat wordt gebruikt, waardoor het kan worden toegepast op een breed scala aan modellen. De techniek maakt gebruik van een "lokale" benadering, waarbij het alleen de omgeving van een specifieke voorspelling analyseert, in plaats van het gehele model. Hierdoor kan LIME worden gebruikt om de besluitvorming van een specifiek model te verklaren en eventuele onregelmatigheden of onverwachte voorspellingen te identificeren. Je kan ook SHAP als techniek gebruiken in een learning-model. Net als LIME is het ontworpen om te begrijpen waarom een model een bepaalde voorspelling doet voor een specifiek item. In tegenstelling tot LIME, gebruikt SHAP de Shapley waardes uit de coöperatieve game theorie om de bijdrage van elk kenmerk aan de voorspelling te bepalen. Deze waardes geven aan hoeveel waarde elk kenmerk toevoegt aan de voorspelling en zorgen voor een meer accurate verklaring van de besluitvorming van een model. SHAP is ook model-agnostisch en kan worden toegepast op verschillende soorten modellen.

Adversarial attack is iets waar je weldegelijk op moet letten bij je training van je ai-model. Dit is namelijk door een bepaalde handling/ hindering een andere situatie creëren. Als je model daar goed is op getraind dan zou de uitkomst niet mogen afwijken van de vorige keren. Een korte beschrijving wat hierbij wordt bedoeld. Je hebt borden herken software in je auto zitten en het moet een stopbord herkennen. Als je nu doormiddel van stickers op het bordt plakt in een ander kleur dan kan het zijn dat de target-attack succes score naar beneden gaat. Omdat het model het bordt niet meer goed herkend. Deze score en robustness moet zo hoog mogelijk zijn zoals in de echte wereld het ook is. Dit kan je ook hebben als je bv een foto hebt van een hond en het model zegt dat deze afbelding geen gond meer is, kan dit te merken hebben men een kleine ruisfilter. Wij zien nog steeds de hond duidelijk maar het model ziet het ruis waardoor hij deze anders herkend. Dit is al eens gebeurd bij een reallife situatie bij tesla, waar ze doormiddel van drie bolletjes op het wegdek een eens specifiek patroon denkt de auto dat hij van lijn moet wijzigen. Dus de auto werd misleid terwijl het geen volle witte lijn is die zegt dat je moet veranderen van vak.

Andere security bedreigingen zijn Membership interference, het beschrijft de mogelijkheid van een aanvaller om aan de hand van een (kleine) hoeveelheid gegevens over een bepaald individu, te bepalen of deze individu deel uitmaakt van een specifieke groep. Attribute inference, het beschrijft de mogelijkheid van een aanvaller om aan de hand van beperkte gegevens over een bepaald individu, te raden welke kenmerken (attributes) deze persoon heeft.

### Conclusie

Bij sommige ethische elementen moet er nog veel duidelijkheid komen van het gebruik van AI-model in het openbaar. Als er accidenten mee gebeuren wie is er verantwoordelijk hiervan? Wat doe teen auto bij een situatie waar hij moet kiezen tussen iemand omverrijden of ergens tegenbotsen? Al deze ethische beslissingen zullen in de toekomst moeten gemaakt worden om duidelijkheid te hebben in de aansprakelijkheid van bepaalde dingen.

Bij het luisteren naar deze sessie heb ik wat bijgeleerd over Het Ai proces, terwijl ik hiervan niet zo veel van wist was het toch een interessante sessie. Vooral het omgaan met de problemen van de ai interesseerde me. En hoe en ai-model alle anders ziet dan wij.