

Bilag 1.

Alt tekst inde i firkantede parentes [...] er til for at uddybe noget fra interviewet eller påpege noget, der skete, som ikke kan uddrages bare ved at læse teksten, som at noget bliver peget på eller refereret til under interviewet. "C/H" er os, der fortager interviewet taget fra vores fornavns initialer (C)asper & (H)enrik. "M" er Mads (Høbye).

5 ----- Interview start -----

C/H: Kan du starte med at definere AI – meget groft

M: Min grundforståelse for AI er, at det er kunstig intelligens. Jeg synes det, der er generativt for mig, er at snakke om AI og ML - altså machine Learning - som to forskellige perspektiver, hvor man kan sige AI er mere den filosofiske perspektiv på det; Det med at computere og systemer kan begynde at tænke selv eller
10 være bevidste om deres handlinger eller blive smartere til at løse komplekse problemer, hvorimod man kan sige Machine Learning er den mere konkrete, datalogiske diskussion om, hvad der kan lade sig gøre i praksis og hvad kan vi med det.

C/H: Så AI er, at maskiner har noget autonomi i at kan tage handlinger selv?

M: Det er mere meta-refleksionen om det; Det er i hvert fald den nemmeste måde for mig at sætte det op
15 på. Jeg er sikker på, at der er tusind andre versioner af den diskussion.

C/H: En anden definition skrevet i en rapport fra et studiepanel lyder: Aktiviteten at gøre maskiner intelligente: Og intelligens som kvaliteten, der tillader en enhed at fungere passende/korrekt og med fremsyn i deres omgivelser. Vil det være en korrekt forklaring af AI?

M: Ja, det er en meget fin og generel beskrivelse af AI og Machine Learning, hvis man blander dem sammen
20 og siger det er det samme.

C/H: Så lad os kigge på nogle tegninger [Se bilagene]. Her har vi tegnet, hvordan vi har forstået den her infrastruktur i det her AI. Det er selvfølgelig groft og der er meget mere hardware involveret. Er denne model mere eller mindre korrekt?

M: Jeg vil sige, at rækkefølgen i jeres model er forkert. Trin 1 burde være kameraet og æblet, 2 en
25 computer og så en match [trin 3]. Den siger jo ikke "Æble". Den siger 70% chance for det er et æble, 20% banan og 10% for det er en pære. Det er jo aldrig 100% korrekt. Den giver en eller anden estimering: Det kunne være det her, men det kunne også være det her. Og nogle gange kan man dreje æblet foran kameraet, og så vil den sige: "pære, pære, pære, æble, æble, banan, banan".

C/H: Køleskabet bruger kameraet til at genkende madvarer internt, så den kan komme med opskrifter. Så
30 for at opklarer, det du siger er, at den siger der er 51% chance for det er et æble, så jeg siger til brugeren det er et æble(?)

M: Man ville normalt lave et threshold, hvor hvis den er 70% sikker på det er et æble, så ser vi det som et æble.

C/H: Hvilke parametre kan den [kunstige intelligens] anvende til at skelne mellem et æble og for eksempel
35 en karton mælk?

M: Det kommer jo an på, hvilke algoritmer du bruger - hvordan den er trænet. Der typisk set en eller anden form for feature extraction med et mellemlid i det her [peger på tegningen, se bilag], hvor den siger... den

trækker nogle features ud: Det kan være alt fra farver til kontraster til mønstre, osv. Og så bruger den dem til at lave genkendelse på. Og så er der forskellige algoritmer med forskellige måder at gøre det på. Nu er vi meget inde i det man kan kalde neurale netværk eller Deep learning systemer, så den gør det på den måde.

C/H: så den kigger på [i reference til et æble]: Det er sådan halvrundt og det har en rød, måske lidt grønlig farve...

M: Sådan noget, ja. Hvis det er en Deep learning algoritme er der faktisk enormt mange lag, de kører ned i; Der er et cluster her af pixels som har høj kontrast, for eksempel. Og så kan jeg se det cluster har nogle venner, der har nogle cluster, der har nogle cluster – så laver den sådan en struktur, som så ender med at være et æble. Den laver mønstre og genkender teksturer på overflader.

C/H: Og du siger det var Deep learning algoritmen. Hvad er alternativet?

M: Altså det er jo bare neurale netværk, men neurale netværk har forskellige algoritmer. Den moderne er den, der hedder Deep learning som så har den mulighed, at den har massere af lag og systemer den bygger det på. Og nogle af dem er også pre-trænet [trænet på forhånd]. Hvor med neurale netværk er det meget klassik - Old school, hvor man giver den en masse æbler og så siger man "det er et æble, det er æble, det er et æble", og så lærer den, at det er et æble. Hvorimod Deep learning, der har man pre-trænet for eksempel teksturer og mønstre og forskellige logikker, og så det eneste man egentlig træner er det på toppen: Om det egentlig er et æble. Der er altså nogle filtre på den.

C/H: Er der en af dem, der er mere effektiv end den anden?

M: Deep learning. Det er det nye hotte, superavancerede. Så når I ser alle de der AI algoritmer, så er det Deep learning.

C/H: Vi læste om nogle programmører, der havde skabt noget [en AI algoritme] der hed YOLO [You Only Look Once]. Det er et Open Source Image recognition software.

M: Ja, YOLO er sådan et pre defineret netværk, hvor den kender en masse objekter. Det er en af de rigtig gode af dem. Men det er det samme man kan her [viser noget, der hedder Teachable Machines] – vi har nogle klassifikationer; Det er sådan det typisk virker. Så vi kan have en klassifikation her, der hedder sten, så kunne vi have en her, der hedder mobil, og så en ny, der hedder ingenting. Altså, hvor der ingenting er foran kameraet. Det ville være en meget traditionel måde at løse det her på; At sige: Jamen vi har de her forskellige klassifikationer – og YOLO-systemet er så bare tusind af sådan nogen – så går jeg ind her, for jeg vil gerne bruge webcammet til at træne. [Mads holder nu de 2 forskellige objekter (sten og mobil) op foran kameraet til deres tilhørende klassifikation] Normalt ville man bruge mange tusind billeder, men her har jeg bare brugt nogle få billeder.

C/H: Så det er ligesom en kontrol, den der hedder ingenting?

M: Ja, det er bare fordi, at der er et scenarie, hvor der ikke er noget foran kameraet. [han illustrerer herefter, at programmet virker]. Så dette er et klassisk system, hvor der er et kamera, der trækker nogle billeder ind, og så klassificere den, hvad de forskellige billeder skal være [forestille], og så trækker den nogle feature ud: Alt fra kontrast til farve til alt muligt. Og så smider man det ind i et neuralt netværk og siger "lær de her ting", og så spytter den de her vurderinger ud [chancen for, at et objekt er en sten, mobil, æble, osv]. Der de her hidden layers i midten [peger på en illustration, se bilag] og så er der et output, der kommer ud.

C/H: Alle de her hvad, undskyld?

M: Hidden layers: det er alle vægtninger, der er interne i systemet. Så den ser et billede, og det der kommer ud i vores tilfælde, det er jo så tre kategorier den skruer på – hvor meget er det et æble, hvor meget en mobil, hvor meget er det ingenting. Vægtningerne er alle de ting mellem inputtet og outputtet, der gør, at den kommer frem til en vurdering. Så er der en propagation algoritme, hvor man siger den skal regne baglæns: For eksempel, jeg vil gerne have det her resultat, og så retter den de her vægtninger til, så de passer. Back-Propagation hedder det. Det der er ved Deep learning det er, at alt det her er meget større og komplekst, hvor nogle af disse elementer er pre-definerede: At den har en forståelse for kontrast og mønstre. Og så det man træner den er kun den sidste del: Altså er det her en pære eller en banan, osv. Så hvis Samsung påstår, de kan det her støder de jo ind i problemer, når folk for eksempel lægger en banan ind i pose og kameraet nu skal genkende det. Og server mæssigt, kan det hele jo køre på en computer, som vi gør lige nu; eller en mobil for den sags skyld. Det er at træne AI, der kræver meget. Så hele den detektor [database] kan jo ligge i køleskabet; Den behøver ikke at gå på nettet for at hente det.

90 ----- Interview slut -----