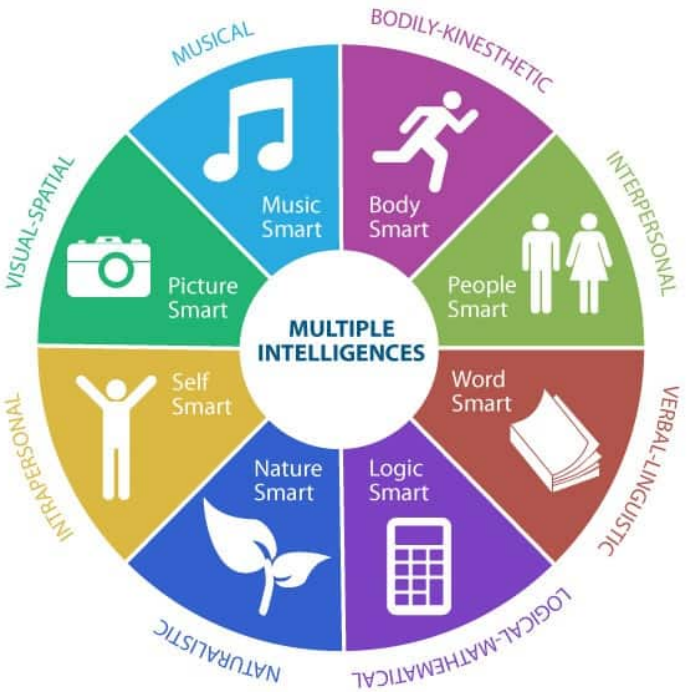
Dit artikel gaat over het **testen van kunstmatige intelligentie** (AI). Ik ga daarbij niet in eerste instantie uit van de techniek achter AI, zoals de meeste presentaties en publicaties doen, maar van het doel van AI. Het doel van AI is om intelligente taken uit te voeren.

Dit artikel bestaat uit drie delen:

1. Wat is intelligentie? Een definitie vanuit inzichten uit de psychologie en een inventarisatie van de belangrijke kenmerken en onderdelen van intelligentie;
2. Hier vanuit leg ik uit hoe de techniek van AI software dit nabootst;
3. Ten slotte ga ik in op het testen van AI: wat zijn de grootste productrisico's in geval van AI en hoe kunnen testers een bijdrage leveren aan het beheersen hiervan?

Wat is intelligentie?

Vraag tien mensen wat intelligentie is en je krijgt tien antwoorden. Dat is niet zo gek, want bijvoorbeeld Gardener (1993) onderscheidt acht verschillende vormen van intelligentie: lichamelijk, interpersoonlijk, talig, logisch, natuur, zelf, beeld, muziek. Iemand kan hoogbegaafd in één intelligentie zijn en laagbegaafd in een andere! Intelligentie is geen aangeboren eigenschap (hersenen zijn zelf niet intelligent), maar het is de kwaliteit van gedrag waarbij begrip van een (deel van) de omgeving of van jezelf wordt ontwikkeld. Intelligentie en lerend vermogen hangen nauw samen. Hersenen spelen natuurlijk een belangrijke rol.



In intelligent gedag zijn de volgende elementen altijd aanwezig:

- Ervaren (verkrijgen van data);
- Filteren (focus op alleen relevante data);
- Ordenen (relaties leggen in de data);
- Conceptualiseren (een model of begrip vormen van een ding, het doel);
- Positioneren/handelen;
- Reflectie, verbinden en variëren van alle vorige stappen.

Te samen vormen deze elementen lerend vermogen.

Kunstmatige intelligentie

In AI software probeert men bovenstaande elementen binnen een systeem te imiteren, zodat software lerend wordt en het zichzelf (deels) aanpast om een taak of begrip te krijgen.

AI software wordt toegepast waar sprake is van veel ongestructureerde en onvoorspelbare data en men niet van te voren weet wat men precies nodig heeft om een taak precies uit te voeren.

Een voorbeeld is beeldherkenning, waarbij men wil dat de software in alle mogelijke foto’s personen of patronen kan identificeren.

AI software bestaat uit:

1. Een eerste analyse van de input. Het resultaat is data die onderworpen kan worden aan de analyse;
2. Een neuraal netwerk: dit bestaat, zoals de hersenen, uit allerlei kleine blokjes code (neuronen) die verbonden zijn met heel veel andere blokjes en die ook wisselende verbindingen aan kan gaan. In het neurale netwerk zijn lagen van verbindingen aangelegd die elk een deel van de analyse uitvoeren. De blokjes zijn geen statische algoritmen zoals we dat kennen uit traditionele software, waarbij elk blokje bij een input aan de hand van een conditie steeds dezelfde output heeft. Elk blokje is in feite leeg, maar heeft parameters die vanuit de interactie met andere blokjes wordt aangepast;
3. De mogelijkheid om het neurale netwerk te corrigeren en aan te geven wat het als geheel moet leveren. Het netwerk wordt een aantal voorbeelden gegeven met de gewenste output (bijvoorbeeld: je voert en groot aantal foto’s in van verschillende dieren en je geeft per foto aan of het een kat of een hond is). Het systeem zal zich dan zo aanpassen dat als het de foto’s ongelabelled zou krijgen, het de foto’s juist zou labellen als kat of hond. Als je daarna 1000 ongelabellede foto’s invoert dan groeit de kans dat het systeem met de juiste labelling komt. De kwaliteit van de labelling neemt toe naarmate je het systeem meer en diverse voorbeelden hebt gevoerd. Dit is machine

learning.

Als het systeem leert aan de hand van vooraf gegeven voorbeelden met de juiste resultaten heet dat supervised learning. Als het systeem zonder voorbeelden zelf komt tot labelling (het ontdekt dat er verschillende groepen dieren zijn, katten en honden) dan heet dit unsupervised learning. Unsupervised learning staat nog in de kinderschoenen.

Testen van kunstmatige intelligentie (AI)

Risico's
Traceerbaarheid: zelfs AI-programmeurs zullen niet weten waar een AI-systeem zijn keuzes op baseert. Ze weten alleen hoe het systeem zich aanpast. Omdat de data waar het systeem mee leert zo veel en divers is, is dit niet meer te traceren.
Reductionisme en overdrijving: als de data waar het systeem mee leert onvolledig, tijdelijk deels beperkt is (gaps, flaws) dan kan bijvoorbeeld labelling onjuist of te simpel zijn.
Ontleren: menselijke intelligentie moet soms ontleren; oude ervaringen kwijt raken of anders duiden (denk aan therapie na trauma's). Dit geldt ook voor AI! Als er eenzijdige data of foutieve data het systeem in gaat (misschien zelfs door hackende andere AI), hoe komt het systeem hier dan van af zonder onterecht andere delen van de geschiedenis kwijt te raken? En kan het systeem ook onterecht vroege ervaringen vergeten? Wordt een simpele huis- tuin- en-keukenkat nog wel herkend als kat als je heel veel exotische katten invoert?
Blind vertrouwen: vanwege bovenstaande risico's zullen mensen blind gaan vertrouwen op AI. We kunnen niet anders, omdat AI een aantal zaken beter kan dan mensen dat kunnen.

Het testen van kunstmatige intelligentie (AI) zal moeten gebeuren door een nieuw soort tester: **de data artist**.

AI-testen gaat over testdata: het aanmaken van testdata, deze kunnen manipuleren, het voorspellen van goede uitkomsten op basis van data. Testers moeten grote datasets beheren en kunnen gebruiken voor regressietests, testrapportages en analyseren van productieresultaten.

Daarnaast is negatief testen nog te noemen: AI-systemen hebben grenzen die ze niet mogen overschrijden, zogenaamde "prime directives". Denk aan drones die geen eigen troepen of burgers mogen raken of een auto die moet kunnen kiezen tussen een hond of een kind om aan te rijden.

Is het mogelijk dat het systeem deze grens overschrijdt? Hier moet de data artist zijn creativiteit en kunde inzetten om situaties te bedenken en het systeem aan te onderwerpen om te kijken of het mogelijk is dat het AI systeem ooit in de fout zou kunnen gaan. Zoals bij alle gecompliceerde testen (zoals E2E-testen) zal de menselijke maat, het daadwerkelijke proces en het gebruik het uitgangspunt moeten blijven.

Gerard Numan

Dit artikel is een samenvatting van de presentatie die Gerard heeft gegeven tijdens de Polteq Conferentie op 7 juni 2018 in het NBC in Nieuwegein.