Advies op het Gebied van Generatieve AI voor SSC-ICT *Door Jitse Goutbeek*

**Aanleiding van dit Document**

Na acht maanden onderzoek binnen het innovatieteam van SSC-ICT, vat ik mijn observaties en voorlopige aanbevelingen met betrekking tot generatieve AI samen in dit document. Dit onderzoek, voortkomend uit de opdracht 'AI en spraakbesturing' (urencode 201202044.002), geaccordeerd door het SPB en gebaseerd op de strategische Service Roadmap van SSC-ICT, was gericht op het verkennen van de mogelijkheden en uitdagingen die generatieve AI biedt voor SSC-ICT. Met de kennis dat mijn rol in dit proces tijdelijk is en dat mijn opvolger dit onderzoek voortzet, bied ik dit document aan als een bron van inzicht en ondersteuning. Dit neemt de vorm aan van een vijftal observaties die leiden tot een advies.

**Observatie 1: veel onzekerheid over de meerwaarde van specifieke toepassingen**

Gesprekken met ambtenaren over de meerwaarde van generatieve AI in hun werkzaamheden leveren een breed scala aan ideeën op. Het blijkt lastig te voorspellen welke ideeën werkelijk potentie hebben. We hebben dusver nog geen toepassingen gevonden waar we[[1]](#footnote-0) resultaten zagen met (consistent) hoge kwaliteit[[2]](#footnote-1), en de manieren waarop de kwaliteit laag was, konden heel verrassend zijn.[[3]](#footnote-2)

Tegelijkertijd zien we een technologisch veld dat zich heel snel ontwikkelt.[[4]](#footnote-3) Het kan heel goed dat wat nu nog niet goed genoeg werkt dat over een paar jaar wel doet[[5]](#footnote-4). Om hier gebruik van te kunnen maken moeten we kunnen vaststellen hoe de kwaliteit van een applicatie zich verhoudt tot de risico’s.

**Observatie 2: er wordt volop geëxperimenteerd binnen de rijksoverheid**

Dit betekent dat we zowel de kwaliteit als risico’s van een generatieve AI toepassing moeten kunnen vaststellen of meten, hier is onderzoek voor nodig.[[6]](#footnote-5) Er zijn veel projecten binnen de rijksoverheid die onderzoeken of een bepaalde generatieve AI toepassingen iets voor hun organisatie kan betekenen.[[7]](#footnote-6) Dit laat zien dat de interesse in deze technologie bij onze klanten groot is, de klanten zelf nog aan het onderzoeken zijn waar de potentie voor hen ligt en soms ook zelf de mensen hebben om applicaties te bouwen.[[8]](#footnote-7)

**Observatie 3: veel onderzoeken lopen tegen een vergelijkbaar probleem aan**

Generatieve AI toepassingen vereisen doorgaans meer rekenkracht dan een overheids laptop kan leveren, en veel onderzoeksprojecten hebben daar geen passend antwoord op.[[9]](#footnote-8) In het bijzonder vereist goed onderzoek dat er veel optimalisaties uitgeprobeerd worden, en steeds gekeken wordt of het voor verbeteringen heeft gezorgd. Dit is een iteratief proces, waarvan de snelheid afhangt van de beschikbare rekenkracht. Op deze manier zorgt weinig rekenkracht er niet alleen voor dat applicaties niet in productie genomen kunnen worden, maar ook dat de kennisopbouw binnen de overheid onnodig langzaam verloopt. Meerdere partijen hebben aangegeven dat ze graag bij dit obstakel geholpen zouden worden.

**Observatie 4: SSC-ICT kan een toekomstige, beheerde dienstverlening aanbieden**

Voor rekenkracht is hardware nodig en de mogelijkheid er op een toegankelijke, verantwoorde en efficiënte manier gebruik van te maken. In tegenstelling tot onze klanten heeft SSC-ICT de contacten met leveranciers[[10]](#footnote-9) om hardware in te kopen en de mensen en expertise (zoals security en hosting[[11]](#footnote-10)) om dat te kunnen leveren[[12]](#footnote-11). De gesprekken die ik gevoerd heb binnen SSC-ICT geven mij het gevoel dat er belangstelling voor deze rol is. Dit kan als beheerde dienst uitgewerkt worden, intern getest en later ook extern geleverd. Veel projecten zijn ook naar ons toegekomen tijdens hun opstartfase voor hulp of kennis, als we zelf een platform aanbieden waar we ervaring mee hebben en doorgaan met kennisopbouw zal het aantal adviesvragen naar verwachting stijgen. Consultancy op maat is ook een dienst die we zouden kunnen leveren.

**Observatie 5: applicaties bouwen heeft meer meerwaarde dan alleen de applicatie zelf.**

Hoewel onze applicatie zelf niet zo goed werkt, hebben we wel een concrete toepassing waarbij we erachter komen wat er nodig is voor je iets in productie kan nemen. We kunnen dit ook gebruiken om een compute platform te testen[[13]](#footnote-12), om kennis op te bouwen die we kunnen delen en om het cruciale werk van validatie verder uit te werken. Voor we iets aanbieden of inkopen moeten we een idee hebben van hoe goed het werkt, en daarvoor moeten we ervaring hebben met dat meten. Bovendien blijft het interessant voor SSC-ICT om verder te onderzoeken welke toepassingen voor ons de meeste potentie hebben. We slaan hiermee twee vliegen in één klap: met meer rekenkracht kunnen we sneller en beter onderzoeken welke toepassingen voor SSC-ICT interessant zijn en dat onderzoek kan gebruikt worden om te onderzoeken op welke manier we het beste de experimenten en toepassingen bij klanten kunnen faciliteren.

**Advies**

Ik acht het voor SSC-ICT verstandig nu in te zetten op de lange termijn. Ben je ervan bewust dat zowel onze kennis van de voor en nadelen als de technologie zelf snel veranderen. Het is nu vaak te vroeg om al een specifieke dienst kant-en-klaar te willen leveren of van de markt in te kopen. Het is heel makkelijk een indrukwekkende demo te laten zien, want dat is vaak gebaseerd op de beste resultaten, maar dat betekent niet dat er niet ook een kans is dat er een heel fout antwoord gegeven wordt met de bijbehorende risico’s - dit is gebaseerd op de slechtste resultaten. We hebben nog niet eens de methoden om die risico’s goed in te schatten, en het gaat nog wel eens fout[[14]](#footnote-13). Geef nu prioriteit aan het opbouwen van kennis en behoud van flexibiliteit in plaats van het eindproduct. Op dit moment kunnen we nog niet snel en goed beoordelen welke potentie en risico’s een markt oplossing voor ons heeft. Door te blijven experimenteren kunnen we steeds beter worden in het beoordelen van de potentie van toepassingen, en niet door hype of angst een oplossing onterecht (niet) gebruiken.

Zet daarom in op dingen die relevant blijven, iedere toepassing heeft rekenkracht nodig en moet getest kunnen worden. Dit kan door bijvoorbeeld een compute platform als beheerde dienst aanbieden, dit willen klanten, hebben we zelf nodig en past in het profiel van SSC-ICT. Hierbij is het belangrijk goed de verschillende opties afwegen voor hoe we het op gebied van hardware in willen richten, het innovatieteam kan een rol spelen dit met concrete toepassingen te testen.

Om op de korte termijn alvast een ‘vervanging voor ChatGPT’ te bieden kunnen we voor een open-source model impact assessments doen[[15]](#footnote-14) en als die positief uitvallen het hosten en aanbieden. Enerzijds (als API) aan ontwikkelaars zodat ze er applicaties mee kunnen testen, anderzijds aan particuliere ambtenaren met een grote melding erbij dat antwoorden niet vertrouwd kunnen worden en het dus alleen geschikt is voor laag risico gebruik.[[16]](#footnote-15)

Als we een compute platform hebben kunnen we ook klanten helpen hier hun ideeën op te testen door een stukje consultancy te leveren. Om dit effectief te doen is het belangrijk dat we er zelf ook gebruik van maken om onze eigen use cases te testen en ervaring op te doen. Sommige klanten zijn zelf al hard bezig en hebben misschien wel meer kennis van generatieve AI dan wij, die hoeven we alleen toegang tot het platform te geven. Anderen moeten een beetje op weg geholpen worden en weer anderen willen misschien liever dat wij bij SSC-ICT zelf het onderzoek op hun verzoek (en kosten) doen.

Hopelijk zorgt dit voor een ecosysteem waarin veel verschillende toepassingen getest worden en we steeds beter zicht krijgen op hoe we die testen goed vorm kunnen geven, waar echt vraag naar is en wat wel en wat niet werkt. Als we vanuit meerdere ministeries zien dat ze interesse hebben in vergelijkbare applicaties die bewezen werken kunnen we er dan als SSC-ICT, in samenwerking met de mensen die de applicaties onderzocht/ontwikkeld hebben, een gerichtere beheerde dienst van maken.

1. Met de huidige stand van de openbaar toegankelijke technologie [↑](#footnote-ref-0)
2. We hebben bijvoorbeeld onderzocht of we met open source tooling onze eigen databases goed door kunnen zoeken (en koppelen aan een taalmodel), maar de huidige tooling is vrij matig en we vonden slechts een minderheid van de relevante documenten goed terug. [↑](#footnote-ref-1)
3. We kregen bijvoorbeeld precies één keer een antwoord in het Argentijns Spaans, zonder aanleiding [↑](#footnote-ref-2)
4. Alle taalmodellen waar ik nu gebruik van maak, bestonden 8 maanden geleden nog niet, en wat wel bestond was veel slechter. [↑](#footnote-ref-3)
5. Het aantal woorden dat je mee kan geven aan een taalmodel is bijvoorbeeld al 50 keer zoveel als een paar jaar terug en de prognose is dat het nog 100 keer meer wordt, dit zou de beperking van een slechte zoekfunctie deels weg kunnen nemen. [↑](#footnote-ref-4)
6. Grote AI labs geven aan dat het vaak lastig te zeggen is of de meetinstrumenten die we hebben wel de relevante dingen meten, voor de overheid geldt dit alleen maar meer (hoe meten we kwaliteit?) [↑](#footnote-ref-5)
7. Ik ken alleen al 8 andere projecten die specifiek kijken naar de potentie van het verbinden van een taalmodel aan de eigen data, en er zullen er meer zijn dan ik ken. [↑](#footnote-ref-6)
8. Het datalab van J&V heeft een heel team dat hiermee bezig is, en meer geld, fte en kennis beschikbaar heeft om applicaties neer te zetten dan er bij SSC-ICT aanwezig is. [↑](#footnote-ref-7)
9. We zijn al 8 maanden bezig met het organiseren van meer rekenkracht, wat nog steeds niet volledig gelukt is. In die tijd is het onderzoek langzamer verlopen en met minder goede resultaten dan het geval was geweest als dit obstakel er niet was geweest. [↑](#footnote-ref-8)
10. Zowel HP als Fujitsu hebben ons een aanbod gedaan om hardware hiervoor [↑](#footnote-ref-9)
11. Meet en houd rekening met het energiegebruik dat hiermee gepaard gaat, anekdotisch zijn er organisaties die wel de hardware ingekocht hadden die vervolgens niet aangesloten kon worden. [↑](#footnote-ref-10)
12. Dit kan op verschillende manieren, bijvoorbeeld binnen ODC of in de cloud. Ik beschrijf hoe dit eruit zou kunnen zien met een lokale oplossing, omdat ik weet dat mensen daarmee bezig zijn. Ik heb echter onvoldoende onderzoek gedaan om te weten of dat beter is dan alternatieven zoals cloud. [↑](#footnote-ref-11)
13. Onze applicatie lijkt vrij veel op applicaties die klanten zouden willen testen, er wordt vaak gebruik gemaakt van deze packages en wij hebben code van andere overgenomen en andere van ons. [↑](#footnote-ref-12)
14. Een applicatie van de New York politie die lijkt op wat wij hebben zei onterecht tegen gebruikers dat illegale praktijken, zoals het houden van fooi van je medewerkers, legaal waren. [↑](#footnote-ref-13)
15. Dit staat op de planning [↑](#footnote-ref-14)
16. Het samenvatten of (in een taal die je kent) vertalen van teksten of helpen brainstormen etc zijn allemaal toepassingen waarbij de gebruiker weet of het antwoord deugt of niet. [↑](#footnote-ref-15)