

# Is AI geavanceerd genoeg om kunstwerken te genereren waarvan de boodschap herkenbaar is uit het dagelijks nieuws?

---

**Dario Bronders.**

Scriptie voorgedragen tot het bekomen van de graad van  
Professionele bachelor in de toegepaste informatica

**Promotor:** Dhr. Manu De Buck

**Co-promotor:** Dhr. Geert van Boven

**Academiejaar:** 2022–2023

**Eerste examenperiode**

**Departement IT en Digitale Innovatie .**

**HO  
GENT**



# Woord vooraf

Deze bachelorproef markeert het hoogtepunt van mijn opleiding Bachelor in Toegepaste Informatica aan de Hogeschool Gent. Mijn fascinatie voor generatieve AI en automatisering, heeft geleid tot het idee om te onderzoeken of AI geavanceerd genoeg is om kunstwerken te genereren die een boodschap uit het dagelijks nieuws overbrengen. De mogelijkheden die deze technologieën bieden in het combineren van kunst, technologie en maatschappij hebben mij enorm gemotiveerd om dit onderzoek uit te voeren.

Tijdens dit onderzoek heb ik kunnen rekenen op de waardevolle begeleiding en expertise van mijn promotor en co-promotor. Mijn oprechte dank gaat uit naar hen voor alle ondersteuning en tijd die zij in mij geïnvesteerd hebben om dit onderzoek tot een goed einde te brengen.

# Samenvatting

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

# Inhoudsopgave

<b>Lijst van figuren</b>	<b>vi</b>
<b>1 Inleiding</b>	<b>1</b>
1.1 Context en achtergrond	2
1.2 Probleemstelling	2
1.3 Onderzoeksvraag	2
1.4 Onderzoeksdoelstelling	3
1.5 Opzet van deze bachelorproef	3
<b>2 Stand van zaken</b>	<b>4</b>
2.1 Generatieve AI	5
2.1.1 GPT	5
2.1.2 GPT4	7
2.2 Webscraping met BeautifulSoup4	8
<b>3 Methodologie</b>	<b>9</b>
<b>4 Conclusie</b>	<b>11</b>
<b>A Onderzoeksvoorstel</b>	<b>13</b>
A.1 Introductie	13
A.2 Literatuurstudie	13
A.2.1 Wat is webscraping?	13
A.2.2 Wat is DALL-E (2)?	14
A.2.3 Wat is Stable Diffusion?	14
A.3 Methodologie	15
A.4 Verwacht resultaten	15
<b>Bibliografie</b>	<b>17</b>

# Lijst van figuren

A.1 voorbeeld: <a href="http://www.hln.be/robots.txt">www.hln.be/robots.txt</a> . . . . .	14
A.2 Voorbeelden van AI-gegenereerde kunstwerken. . . . .	16

# 1

## Inleiding

In een wereld waarin technologie voortdurend evolueert, rijst de vraag of generatieve kunstmatige intelligentie (AI) al geavanceerd genoeg is om betekenisvolle creaties te genereren die een boodschap overbrengen op basis van dagelijks nieuws. Deze bachelorproef onderzoekt de huidige stand van generatieve AI en beoordeelt de effectiviteit en kwaliteit van AI-generaties bij het communiceren van nieuwsgerelateerde boodschappen op unieke en boeiende wijze met het publiek. Door de combinatie van geavanceerde AI-technieken en automatisering van webscraping, verkennen we de grenzen van de interactie tussen technologie en informatievoorziening, en bepalen we of generatieve AI al voldoende ontwikkeld is om deze uitdaging aan te gaan.

## 1.1. Context en achtergrond

In de afgelopen jaren heeft kunstmatige intelligentie (AI) aanzienlijke vooruitgang geboekt, met name op het gebied van generatieve modellen. Deze vooruitgang heeft geresulteerd in geavanceerde systemen die in staat zijn om zelfstandig creatieve output te genereren. Het potentieel van deze technologieën om kunst en communicatie te transformeren is enorm. Hieronder volgt een overzicht van de belangrijkste ontwikkelingen en trends op het gebied van generatieve AI.

De relevantie van dit onderzoek ligt in de mogelijkheid om een brug te slaan tussen kunst, technologie en maatschappij. Het kan leiden tot nieuwe manieren van nieuwscommunicatie en artistieke expressie. De methodologie omvat het gebruik van AI-algoritmen en webscraping om kunstwerken te genereren op basis van actuele gebeurtenissen.

## 1.2. Probleemstelling

Het is onduidelijk of AI geavanceerd genoeg is om kunstwerken te creëren die een nieuwsgelateerde boodschap effectief overbrengen. Dit onderzoek zal deze vraag beantwoorden en de praktische toepassingen van dergelijke technologieën verkennen voor een doelgroep van kunstenaars, nieuwsorganisaties en ontwikkelaars.

## 1.3. Onderzoeksvraag

Kan AI geavanceerde kunstwerken genereren die een boodschap uit het dagelijks nieuws effectief overbrengen, en zo ja, hoe kunnen deze technologieën worden toegepast in de context van kunst, communicatie en maatschappij?

Deelvragen zijn onder andere:

- Welke generatieve AI-technieken zijn het meest geschikt voor het creëren van nieuwsgelateerde kunstwerken?
- Hoe kunnen webscraping-technieken gebruiken om nieuwsartikelen te kunnen scrapen?
- In welke mate kan AI effectief de essentie van een nieuwsverhaal vastleggen en vertalen naar visuele kunst?
- Hoe kunnen AI-gegenereerde kunstwerken worden geëvalueerd op hun vermogen om een nieuwsboodschap over te brengen?



## **1.4. Onderzoeksdoelstelling**

Het doel van deze bachelorproef is om de mogelijkheden van generatieve AI te onderzoeken bij het genereren van kunstwerken die nieuwsgelerateerde boodschappen communiceren. Hierbij zal er specifiek worden onderzocht of generatieve AI al geavanceerd genoeg is om dergelijke kunstwerken te creëren en welke technieken hiervoor het meest geschikt zijn.

Deze bachelorproef zal zich richten op kunstenaars, onderzoekers en bedrijven die geïnteresseerd zijn in de mogelijkheden van generatieve AI voor het creëren van betekenisvolle kunstwerken. Het uiteindelijke doel is om een aanbeveling te doen over de haalbaarheid en effectiviteit van generatieve AI voor het communiceren van nieuwsgelerateerde boodschappen door middel van kunst.

## **1.5. Opzet van deze bachelorproef**

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

In Hoofdstuk 4, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

# 2

## Stand van zaken

Het genereren van kunstwerken op basis van tekstuele input is een relatief nieuw en innovatief onderzoeksgebied dat de afgelopen jaren aanzienlijke vooruitgang heeft geboekt. Er zijn verschillende technieken en methoden ontwikkeld om deze kunstwerken te creëren, variërend van traditionele machinale vertaling en samenvatting tot meer geavanceerde deep learning-algoritmen.

In deze literatuurstudie zullen we een uitgebreid overzicht geven van de huidige stand van zaken op het gebied van het genereren van kunstwerken op basis van tekstuele input. We bespreken de belangrijkste technieken en methoden die momenteel worden gebruikt en onderzoeken de resultaten die zijn behaald in verschillende toepassingsgebieden. Daarnaast bespreken we ook de uitdagingen en beperkingen van deze technologie, evenals mogelijke toekomstige ontwikkelingen op dit gebied.

We zullen ons in het bijzonder richten op het gebruik van webscraping als een methode om tekstuele input te verzamelen voor het genereren van kunstwerken. We zullen bespreken hoe deze methode werkt, welke uitdagingen er zijn bij het gebruik ervan en hoe deze technologie kan worden gebruikt om informatie te verzamelen.

## **2.1. Generatieve AI**

Generatieve AI is een subdomein van kunstmatige intelligentie dat zich richt op het creëren van nieuwe, unieke content op basis van bestaande gegevens. Het maakt gebruik van geavanceerde technieken zoals deep learning en neurale netwerken om patronen en relaties in de data te identificeren en te leren. Dit stelt de AI in staat om innovatieve en creatieve oplossingen te genereren voor verschillende toepassingen, zoals het genereren van kunstwerken op basis van tekstuele input (Sonix.ai, [2022](#))

Een voorbeeld van een toepassing van generatieve AI is het RobotReporter-project, uitgevoerd aan de Hogeschool Utrecht. Dit project onderzoekt de mogelijkheden van generatieve AI-systemen om journalistiek werk te automatiseren. Het project richt zich op het gebruik van webscraping en natural language processing (NLP) om informatie te verzamelen en te verwerken, en vervolgens met behulp van generatieve AI-algoritmen nieuwe content te creëren (Utrecht, [2021](#)). Dit kan leiden tot efficiëntere en snellere nieuwsproductie.

Generatieve AI heeft niet alleen invloed op de journalistieke sector, maar ook op andere beroepen. Een artikel van Harvard Business Review bespreekt de impact van generatieve AI op klantenserviceberoepen (Review, [2023](#)). In plaats van banen te vervangen, wordt betoogd dat generatieve AI klantenservicemedewerkers kan ondersteunen en hun werk kan verbeteren. Door geautomatiseerde systemen te gebruiken om routineuze taken te voltooien, kunnen medewerkers zich concentreren op meer complexe en empathische aspecten van hun werk.

In de afgelopen jaren is generatieve AI een veelbelovend onderzoeksgebied geworden, met tal van toepassingen en mogelijkheden. Desondanks zijn er nog steeds uitdagingen en beperkingen die moeten worden aangepakt, zoals de ethische aspecten en de benodigde rekenkracht. Toekomstig onderzoek en ontwikkeling zullen waarschijnlijk leiden tot nieuwe doorbraken en toepassingen op het gebied van generatieve AI.

### **2.1.1. GPT**

Generative Pre-trained Transformer (GPT) is een reeks taalmodellen ontwikkeld door OpenAI. Deze modellen zijn ontworpen om menselijke taal te begrijpen en te genereren met een hoge mate van nauwkeurigheid en coherentie. De reeks GPT-modellen omvat GPT-3, GPT-3.5 en GPT-4. GPT-modellen zijn relevant voor de vraag of AI in staat is kunstwerken te genereren die niet van echt te onderscheiden zijn, vanwege hun prestaties op het gebied van natuurlijke taalverwerking en tekstgeneratie.

In de volgende subsecties zullen we de kenmerken en prestaties van elk model bespreken, evenals hun bijdragen aan het veld van kunstmatige intelligentie en hun potentieel om kunstwerken te helpen genereren.

### GPT-3

GPT-3, is een transformer-gebaseerd taalmodel met 175 miljard parameters. Het model heeft interessante prestaties geleverd op het gebied van natuurlijke taalverwerking en kan tekst genereren. (Metz, 2020). GPT-3 kan worden gebruikt om teksten te genereren, zoals nieuwsartikelen, verhalen, gedichten en zelfs code. Hoewel het model niet perfect is en soms onjuiste of irrelevante informatie kan genereren, heeft het de aandacht getrokken van onderzoekers, technologiebedrijven en het grote publiek vanwege de veelzijdigheid en de kwaliteit van de gegenereerde tekst (Wikipedia, 2021).

GPT-3 is gelanceerd door OpenAI en heeft een ongekennde groei doorgemaakt, met een recordaantal gebruikers in korte tijd. Binnen slechts twee maanden na lancering bereikte het platform 100 miljoen gebruikers, waarmee het de snelst groeiende platform ooit is. (Hu, 2023). Deze snelle groei en brede acceptatie zijn indicatief voor de potentie en relevantie van GPT-3 voor deze paper.

### GPT-3.5

GPT-3.5, een geüpgradede versie van het GPT-3-model, biedt geavanceerdere taalbegrip- en generatiecapaciteiten in vergelijking met zijn voorganger. De verbeteringen in GPT-3.5 stellen het model in staat om een breder scala aan complexe taken uit te voeren en betere resultaten te leveren in verschillende toepassingsgebieden, zoals tekstclassificatie, sentimentanalyse, tekstgeneratie en contextueel begrip (Junjie, 2023; Nappier, 2023).

(Nappier, 2023) bespreken de technische aspecten van GPT-3.5 en de belangrijkste verschillen tussen GPT-3 en GPT-3.5. Een van de opmerkelijke verbeteringen in GPT-3.5 is het vermogen om context beter te begrijpen en relevante tekst te genereren op basis van die context. Dit is vooral belangrijk in taken waarbij contextuele informatie cruciaal is voor het begrijpen en genereren van geschikte antwoorden of voortzettingen van een tekst.

Een ander belangrijk aspect van GPT-3.5 is de verbeterde efficiëntie en prestaties. GPT-3.5 maakt gebruik van geoptimaliseerde architecturen en trainingsmethoden om de modelgrootte en computatievereisten te verminderen zonder concessies te doen aan de prestaties. Dit maakt het model toegankelijker en bruikbaar voor een breder scala aan toepassingen en apparaten (Nappier, 2023).

(Junjie, 2023) verkennen de toepassingsmogelijkheden van GPT-3.5 en presenteren verschillende casestudy's die de prestaties en effectiviteit van het model aantonen. Ze onderzoeken het gebruik van GPT-3.5 in diverse domeinen, zoals het genereren van nieuwsartikelen, samenvatten van teksten, het beantwoorden van vragen op basis van tekst en het genereren van code.

Ondanks de goede prestaties van GPT-3.5, zijn er nog steeds uitdagingen en beperkingen die moeten worden aangepakt. In (Junjie, 2023) worden enkele van deze problemen besproken, zoals het genereren van irrelevante of onjuiste informatie (hallucineren), gevoeligheid voor vooroordelen in de trainingsdata en het vermogen om lange teksten te genereren zonder af te dwalen van het onderwerp. Toekomstig onderzoek zou zich kunnen richten op het verbeteren van deze aspecten en het ontwikkelen van methoden om de modelprestaties verder te verfijnen.

### **GPT4**

GPT-4, de opvolger van GPT-3.5, is een nog geavanceerder transformer-gebaseerd taalmodel dat verder bouwt op de successen en verbeteringen van zijn voorgangers. GPT-4 biedt aanzienlijke verbeteringen op het gebied van natuurlijke taalverwerking, tekstgeneratie en begrip, evenals efficiëntie en bruikbaarheid (Microsoft, 2023; OpenAI, 2023).

Een belangrijk kenmerk van GPT-4 is het vermogen om nog nauwkeuriger en coherenter menselijke taal te genereren en te begrijpen. Hierdoor kan GPT-4 beter presteren in verschillende toepassingsgebieden, zoals tekstclassificatie, sentimentanalyse, tekstgeneratie, contextueel begrip en vele anderen (Research, 2023).

(Junjie, 2023) onderzoeken de technische aspecten en verbeteringen van GPT-4 ten opzichte van GPT-3.5. Ze benadrukken de geoptimaliseerde architecturen en trainingsmethoden die in GPT-4 zijn geïmplementeerd, waardoor het model betere prestaties kan leveren zonder dat dit ten koste gaat van de efficiëntie. Deze verbeteringen maken GPT-4 toegankelijker en bruikbaar voor een breder scala aan toepassingen en apparaten.

Binnen het onderzoek van Microsoft (Research, 2023) toont men aan dat GPT-4 vormen van algemene intelligentie vertoont. Dit blijkt uit de kerncapaciteiten, zoals redeneren, creativiteit en deductie, expertise op verschillende onderwerpen, en de verscheidenheid aan taken die het kan uitvoeren. Hoewel er nog veel werk te doen is om een volledige AGI (Artificial General Intelligence) te creëren, wordt benadrukt dat het definiëren van intelligentie, AI en AGI complex en controversieel is en dat er geen definitieve definitie bestaat. Het onderzoek suggereert dat toekomstig werk

op het gebied van GPT-4 en vergelijkbare systemen zich kan richten op het verkennen van nieuwe toepassingen en domeinen en het begrijpen van de mechanismen en principes die aan hun intelligentie ten grondslag liggen.

Naast de prestaties van GPT-4 hebben Elon Musk en andere experts opgeroepen tot een tijdelijke stop van de ontwikkeling van AI-systemen die GPT-4 kunnen overtreffen, vanwege potentiële risico's en onvoorziene gevolgen die dergelijke geavanceerde systemen met zich mee kunnen brengen (Reuters, [2023](#)).

## 2.2. Webscraping met BeautifulSoup4

Webscraping is het proces van het extraheren van informatie uit websites door de onderliggende HTML- en CSS-code te analyseren. BeautifulSoup4 (BS4) is een populaire Python-bibliotheek voor webscraping die het eenvoudig maakt om gegevens van webpagina's te verkrijgen en te verwerken (Chunmei Zheng, [2015](#); Daniel, [2014](#)).

Webscraping kunnen we toepassen in drie basisstappen: (1) het ophalen van de HTML-inhoud van een webpagina, (2) het parsen van de HTML met BS4 om een parse tree te genereren, en (3) het navigeren en extraheren van gegevens uit de parse tree met behulp van BS4-methoden en -functies. (Daniel, [2014](#))

Kiran et al. (Chunmei Zheng, [2015](#)) bieden een uitgebreide handleiding voor webscraping in Python met BS4. Ze bespreken de basisprincipes van webscraping en de belangrijkste functies van BS4, zoals het zoeken naar elementen op basis van tags, attributen en tekstinhoud, en het manipuleren van de parse tree om gegevens te extraheren en te verwerken.

De officiële documentatie van BeautifulSoup is een waardevolle bron van informatie en richtlijnen voor het gebruik van de BS4-bibliotheek. Het bevat gedetailleerde uitleg over de verschillende functionaliteiten, methoden en voorbeeldcode voor het uitvoeren van webscrapingtaken. De documentatie is beschikbaar op de volgende website: (Richardson, [2023](#)).

# 3

## Methodologie

Etiam pede massa, dapibus vitae, rhoncus in, placerat posuere, odio. Vestibulum luctus commodo lacus. Morbi lacus dui, tempor sed, euismod eget, condimentum at, tortor. Phasellus aliquet odio ac lacus tempor faucibus. Praesent sed sem. Praesent iaculis. Cras rhoncus tellus sed justo ullamcorper sagittis. Donec quis orci. Sed ut tortor quis tellus euismod tincidunt. Suspendisse congue nisl eu elit. Aliquam tortor diam, tempus id, tristique eget, sodales vel, nulla. Praesent tellus mi, condimentum sed, viverra at, consectetur quis, lectus. In auctor vehicula orci. Sed pede sapien, euismod in, suscipit in, pharetra placerat, metus. Vivamus commodo dui non odio. Donec et felis.

Etiam suscipit aliquam arcu. Aliquam sit amet est ac purus bibendum congue. Sed in eros. Morbi non orci. Pellentesque mattis lacinia elit. Fusce molestie velit in ligula. Nullam et orci vitae nibh vulputate auctor. Aliquam eget purus. Nulla auctor wisi sed ipsum. Morbi porttitor tellus ac enim. Fusce ornare. Proin ipsum enim, tincidunt in, ornare venenatis, molestie a, augue. Donec vel pede in lacus sagittis porta. Sed hendrerit ipsum quis nisl. Suspendisse quis massa ac nibh pretium cursus. Sed sodales. Nam eu neque quis pede dignissim ornare. Maecenas eu purus ac urna tincidunt congue.

Donec et nisl id sapien blandit mattis. Aenean dictum odio sit amet risus. Morbi purus. Nulla a est sit amet purus venenatis iaculis. Vivamus viverra purus vel magna. Donec in justo sed odio malesuada dapibus. Nunc ultrices aliquam nunc. Vivamus facilisis pellentesque velit. Nulla nunc velit, vulputate dapibus, vulputate id, mattis ac, justo. Nam mattis elit dapibus purus. Quisque enim risus, congue non, elementum ut, mattis quis, sem. Quisque elit.

Maecenas non massa. Vestibulum pharetra nulla at lorem. Duis quis quam id lacus dapibus interdum. Nulla lorem. Donec ut ante quis dolor bibendum condimentum. Etiam egestas tortor vitae lacus. Praesent cursus. Mauris bibendum pede at elit. Morbi et felis a lectus interdum facilisis. Sed suscipit gravida turpis. Nulla at

lectus. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Praesent nonummy luctus nibh. Proin turpis nunc, congue eu, egestas ut, fringilla at, tellus. In hac habitasse platea dictumst.

Vivamus eu tellus sed tellus consequat suscipit. Nam orci orci, malesuada id, gravida nec, ultricies vitae, erat. Donec risus turpis, luctus sit amet, interdum quis, porta sed, ipsum. Suspendisse condimentum, tortor at egestas posuere, neque metus tempor orci, et tincidunt urna nunc a purus. Sed facilisis blandit tellus. Nunc risus sem, suscipit nec, eleifend quis, cursus quis, libero. Curabitur et dolor. Sed vitae sem. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Maecenas ante. Duis ullamcorper enim. Donec tristique enim eu leo. Nullam molestie elit eu dolor. Nullam bibendum, turpis vitae tristique gravida, quam sapien tempor lectus, quis pretium tellus purus ac quam. Nulla facilisi.



# 4

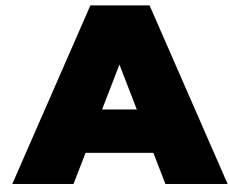
## Conclusie

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

Quisque enim. Proin velit neque, tristique eu, eleifend eget, vestibulum nec, lacus. Vivamus odio. Duis odio urna, vehicula in, elementum aliquam, aliquet laoreet, tellus. Sed velit. Sed vel mi ac elit aliquet interdum. Etiam sapien neque, convallis et, aliquet vel, auctor non, arcu. Aliquam suscipit aliquam lectus. Proin tincidunt magna sed wisi. Integer blandit lacus ut lorem. Sed luctus justo sed enim.

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetur libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem. Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.



# Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

## **A.1. Introductie**

Traditionele nieuwsbronnen streven al geruime tijd naar objectiviteit en feitelijkheid, om informatie vanuit eenzelfde perspectief en met een vergelijkbare boodschap over te brengen. Door het hoogtepunt van de dag te abstraheren tot een uniek AI-gegenereerd kunstwerk, kan er worden geëxperimenteerd met de grenzen van de menselijke perceptie en kunst.

Mijn toegepaste onderzoek richt zich op het ontwikkelen van een applicatie die kunstwerken genereert op basis van het dagelijkse hoogtepunt, waarvoor gegevens zullen worden verzameld door het 'scrapen' van nieuwswebsites en socialmediaplatformen. Deze zal een kunstwerk genereren met behulp van één of meerdere deep learning modellen.

Het doel is om te onderzoeken of AI kunstwerken kan genereren die een boodschap van het dagelijks nieuws kan overbrengen. Dit zal gebeuren met behulp van een turing test die zal bevestigen of het kunstwerk al dan niet enige emotie of boodschap overbrengt met enige relevantie tot het nieuws dat die dag gebeurt.

Dit onderzoek kan interessante inzichten opleveren in de relatie tussen kunst, technologie en de samenleving.

## **A.2. Literatuurstudie**

### **A.2.1. Wat is webscraping?**

Webscraping is een term die gebruikt wordt voor het extraheren van inhoud van websites om het te importeren in lokale opslag zoals een database of CSV bestand.

(Salem & Mazzara, 2020)

Websites kunnen ervoor kiezen om een *robots.txt* (A.1) in de root van hun filesystem te plaatsen. Binnen deze tekstfile kunnen ze beschrijven welke routes gescraped mogen worden. (Google, 2022)

```
# Alle auteurs-, nabuurt- en databankrechten die op de inhoud en opmaak van de DPG Media websites
# en DPG Media apps rusten, worden door DPG Media BV uitdrukkelijk voorbehouden. De inhoud van de
# DPG Media websites en apps is uitsluitend voor persoonlijk, niet-commercieel gebruik en het is
# niet toegestaan om gegevens van de website of uit de apps door middel van screen scraping
# (of een andere automatiseerde werkwijze) te vergaren.
# Zie ook de Gebruikersvoorwaarden van DPG Media op www.dpgmedia.be/gebruikersvoorwaarden

# All copyrights, neighbouring rights and database rights in the content and layout of the
# DPG Media websites and DPG Media apps are explicitly reserved by DPG Media BV. The content of the DPG Media
# websites and DPG Media apps is for personal, non-commercial use only and it is not allowed to
# collect data from the website or from the apps by means of screen scraping (or any other
# automated method).
# See also the terms of use of DPG Media at www.dpgmedia.be/gebruikersvoorwaarden

# Tell robots which pages are not very interesting
User-agent: *
Disallow: /*webview
Disallow: /auth
Disallow: /*widget*
Disallow: /*?tag=
Disallow: /*?abo_type=
Disallow: /*?site_source=
Disallow: /zoeken?query=
Disallow: /inloggen*
# Tell robots not to crawl redirect urls
Disallow: /*?redirect_url=*

User-agent: Twitterbot
Allow: /

Sitemap: https://www.hln.be/sitemap.xml
Sitemap: https://www.hln.be/sitemap-news.xml
```

**Figuur (A.1)**

voorbeeld: [www.hln.be/robots.txt](https://www.hln.be/robots.txt)

### A.2.2. Wat is DALL-E (2)?

DALL-E is een AI software ontwikkeld door openAI dat beelden creëert uit tekstuele beschrijvingen, ook wel *prompts* genoemd. Het gebruikt een versie met 12 miljard parameters van het GPT-3 Transformer-model om natuurlijke taalinput te interpreteren en overeenkomstige beelden te genereren. In april 2022 heeft OpenAI DALL-E 2 gelanceerd, ontwikkeld om meer realistische foto's met hogere resolutie te kunnen genereren.

([nl.wikipedia.org](https://nl.wikipedia.org), 2022) (en.[wikipedia.org](https://wikipedia.org), 2022a)

DALL-E 2 is bovendien getraind met behulp van 650 miljoen tekstinputs gescraped van het internet. (Borji, 2022)

DALL-E 2 is niet open source maar kun je gebruiken aan de hand van de openAI API.

### A.2.3. Wat is Stable Diffusion?

Stable Diffusion is een deep learning, tekst-naar-beeld model uitgebracht in 2022. In tegenstelling tot DALL-E (2) is Stable Diffusion getraind aan de hand van een diepe generatieve neurale netwerk. (en.[wikipedia.org](https://wikipedia.org), 2022b)

Stable Diffusion is open source en kun je lokaal draaien op een computer met een GPU.

## A.3. Methodologie

### Inleiding

Het toegepaste onderzoek start op 2 maart 2023 en eindigt voor 28 mei 2023.

#### Fase 1: Realiseren van een scraper

Om de data te bekomen van de verschillende soorten websites of social-media platformen zal er een web scraper worden gemaakt. Deze scraper zal ontwikkeld worden in python met behulp van een externe library *BeautifulSoup*.

#### Fase 2: Data verwerken en analyseren

Tijdens de tweede fase zullen we onderzoeken op welke manier we de bekomen data uit voorgaande fase kunnen analyseren en sorteren.

Het zal belangrijk zijn om rekening te houden met de volgende vragen:

- Wat zijn de te extraheren kernzaken?
- Wat is het sentiment van de dag?
- Welke topic komt het vaakst voor?
- Op basis van welke gegevens kunnen we de artikels sorteren?

Nadat er een gepaste methode wordt gevonden om dit te realiseren, zal deze ook geïmplementeerd worden. Op deze manier kunnen we steeds het belangrijkste artikel van de dag eruit halen.

#### Fase 3: Kunstwerk genereren

Nu dat we weten uit de vorige fase wat het hoogtepunt van de dag was. Kunnen we hierop een kunstwerk laten genereren.

Hiervoor zal er gebruik gemaakt worden van een of meerdere deep learning modellen DALL-E 2 en/of Stable Diffusion die de kerntekst van een artikel zal omvormen tot een foto.

#### Fase 4: Turing test

Tijdens de laatste fase van dit onderzoek willen we bepalen of de AI-gegenereerde kunstwerken van DALL-E 2 en/of Stable Diffusion herkenbaar zijn uit het dagelijks nieuws. We zullen hiervoor een Turing test uitvoeren dit zal beoordelen.

## A.4. Verwacht resultaten

Het doel van het project is om een goed functionerende applicatie te ontwikkelen die dagelijks een kunstwerk kan genereren op basis van het hoogtepunt van de dag. Bijvoorbeeld, wanneer Marokko won van België tijdens de WK, kan het hoogtepunt in België *'Riots in Brussels after soccer game, painting'* zijn. Op basis van deze tekstinput kunnen enkele voorbeelden gegenereerd worden met behulp van

DALL-E 2. Deze voorbeelden worden weergegeven in de onderstaande afbeeldingen.



**Figuur (A.2)**

Voorbeelden van AI-gegenereerde kunstwerken.

Daarnaast zal uit de turing test blijken of het kunstwerk enige boodschap overbrengt van het nieuws. Een voorbeeld van het resultaat van het project zou bijvoorbeeld kunnen zijn dat op een dag het hoogtepunt een artikel is over een nieuwe doorbraak in de gezondheidszorg. De applicatie zal dan op basis van de tekst van dat artikel een kunstwerk genereren dat hierbij past. In de verwachte resultaten zou dan bijvoorbeeld een afbeelding kunnen worden opgenomen van het gegenereerde kunstwerk op basis van enkele hoogtepunten van de dag. Als de gebruiker hier enige nieuws in kan herkennen en duidelijk kan verwoorden. Dan is dit een geslaagde test.

# Bibliografie

- Borji, A. (2022). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. *Computer Vision and Pattern Recognition*.
- Chunmei Zheng, Z. P., Guomei He. (2015). A Study of Web Information Extraction Technology Based on Beautiful Soup. *Journal of Computers*. <http://www.jcomputers.us/vol10/jcp1006-03.pdf>
- Daniel, H., Analia. (2014). Web scraping technologies in an API world. *Briefings in Bioinformatics*, 15(5), 788–797. <https://academic.oup.com/bib/article/15/5/788/2422275>
- en.wikipedia.org. (2022a). *DALL-E*. Wikipedia. <https://en.wikipedia.org/wiki/DALL-E>
- en.wikipedia.org. (2022b). *Stable Diffusion*. Wikipedia. [https://en.wikipedia.org/wiki/Stable\\_Diffusion](https://en.wikipedia.org/wiki/Stable_Diffusion)
- Google. (2022). *Introduction to robots.txt*. Google. <https://developers.google.com/search/docs/crawling-indexing/robots/intro>
- Hu, K. (2023). *ChatGPT sets record for fastest-growing user base: analyst note*. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- Junjie, N., Xuanning. (2023). A Comprehensive Capability Analysis of GPT-3 and GPT-3.5 Series Models. <https://arxiv.org/ftp/arxiv/papers/2303/2303.10420.pdf>
- Metz, C. (2020). *Meet GPT-3. It Has Learned to Code (and Blog and Argue)*. <https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>
- Microsoft, O. (2023). Capabilities of GPT-4 on Medical Challenge Problems. <https://arxiv.org/pdf/2303.13375.pdf>
- Nappier, I. (2023). *GPT-4 to GPT-3.5: 'Hold My Scalpel' – A Look at the Competency of OpenAI's GPT on the Plastic Surgery In-Service Training Exam* (masterscriptie). <https://arxiv.org/ftp/arxiv/papers/2304/2304.01503.pdf>
- nl.wikipedia.org. (2022). *DALL-E*. nl.wikipedia.org: DALL-E. <https://nl.wikipedia.org/wiki/DALL-E>
- OpenAI. (2023). *GPT-4*. OpenAI. <https://openai.com/product/gpt-4/>
- Research, M. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. <https://arxiv.org/pdf/2303.12712.pdf>
- Reuters. (2023). *Elon Musk and others urge AI pause, citing 'risks to society'*. Reuters. <https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/>

- Review, H. B. (2023). *Generative AI Will Enhance, Not Erase, Customer Service Jobs*. Harvard Business Review. <https://hbr.org/2023/03/generative-ai-will-enhance-not-erase-customer-service-jobs>
- Richardson, L. (2023). *Beautiful Soup Documentation*. Crummy. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- Salem, H., & Mazzara, M. (2020). Pattern Matching-based scraping of news websites. *Journal of Physics: Conference Series*, 1694, 6.
- Sonix.ai. (2022). *Wat is generatief AI?* Sonix.ai. <https://sonix.ai/resources/nl/wat-is-generatief-ai/>
- Utrecht, H. (2021). *RobotReporter: Onderzoek naar generatieve AI-systemen*. Hogeschool Utrecht. <https://www.hu.nl/onderzoek/projecten/robotreporter-onderzoek-naar-generatieve-ai-systemen>
- Wikipedia. (2021). *GPT-3*. Wikipedia. <https://en.wikipedia.org/wiki/GPT-3>