

Perditum in translatione:

Culturele overdracht in de handen van AI

Een vergelijkende MQM-studie naar culturele verschuivingen in LLM-vertalingen van Latijnse epigrammen.

Kobe Geerts

Masterproef aangeboden binnen de opleiding
master in de taal- en letterkunde

Promotor: prof. Lieve De Wachter
Co-promotor: dr. Jordi Heeren

Academiejaar 2025-2026

105.752 tekens



*Copyright informatie:
studiewerk van een student in het kader van de academische
opleiding en examenbeoordeling.
Na deze beoordeling vond geen correctie plaats
van het studiewerk.*

Ik verklaar me akkoord met de code of conduct van de Faculteit Letteren
voor geloofwaardig auteurschap.

Dankwoord

Het schrijven van deze masterproef was een intensief leerproces dat ik onmogelijk alleen had kunnen voltooien. Het brengt me dan ook een enorme voldoening om mijn dankbaarheid te kunnen uiten aan iedereen die me hierbij gesteund heeft. Allereerst ben ik onzegbaar veel dank verschuldigd aan mijn promotor en copromotor, Lieve De Wachter en Jordi Heeren, om mijn thesis te begeleiden. Van begin tot einde kon ik op hun steun rekenen tijdens het brainstormen, bij mijn vragen en in de vorm van hulpvaardige feedback. Zij stuurden deze thesis in de juiste richting en deden dit met zodanig veel enthousiasme dat ik er nog steeds gelukkig van word.

Ten tweede wil ik graag Stef Accou bedanken voor zijn waardevol advies en de tijd die hij maakte om zijn goede raad met me te delen. In hetzelfde verband bedank ik ook graag de docenten, die ik gecontacteerd heb tijdens onder meer de corpusseselectie, voor hun inspiratie. Verder wil ik mijn dankbaarheid onder woorden brengen aan mijn dierbare medestudenten, die steeds bereid waren om vragen en twijfels op te lossen. Ik heb aan hen ware vrienden overgehouden. Daarnaast wil ik iedereen bedanken die heeft geholpen in mijn studie van Klassieke Talen, waaronder mijn goede vrienden van het middelbaar en mijn geëngageerde leerkrachten, zonder wie ik nooit tot dit punt zou zijn geraakt.

Ook mijn vriendin, die onvoorwaardelijke steun en liefde gaf, wens ik uit de grond van mijn hart te bedanken. De waarde van haar geduld en motiverende woorden voor deze masterproef is niet te onderschatten. Ten slotte ben ik ontzettend dankbaar voor de steun en hulp van mijn ouders. Zij boden een luisterend oor tijdens moeilijke momenten, gaven onmisbaar advies en maakten opofferingen zoals enkel ouders dat kunnen.

Bedankt iedereen!

Samenvatting

De snelle opkomst van Large Language Models (LLM's) roept vragen op over hun omgang met *low-resource*-talen en het daarin vervatte culturele gedachtegoed. Deze studie onderzoekt hoe *general-purpose* LLM's culturele concepten uit het Latijn vertalen, met de epigrammen van Martialis als casus. Om de aard van deze vertaalverschuivingen te kwantificeren, is een handmatige annotatie uitgevoerd volgens het *Multidimensional Quality Metrics* (MQM)-framework, gefaciliteerd door aangepaste categorieën en een speciaal hiertoe ontwikkelde applicatie. Hierbij worden de vertalingen door de modellen Gemini 3.1 Pro, M365 Copilot Chat en Kimi-K2.6 vergeleken met de brontekst en twee menselijke referentievertalingen. Uit de analyse blijkt enerzijds een duidelijke tendens naar letterlijkere, generiekere vertalingen die culturele nuance missen, anderzijds een assimilerende vertaalstrategie, vermoedelijk gedreven door een Engelstalige bias in de trainingsdata. Bij de modelvergelijking produceerde Gemini 3.1 Pro significant minder verschuivingen dan de andere twee modellen. Ten slotte tonen de resultaten aan dat langere bronteksten resulteren in een proportioneel hoger aantal verschuivingen per vers.

Abstract

The rapid rise of Large Language Models (LLMs) raises questions regarding their handling of low-resource languages and the cultural heritage embedded within them. This study investigates how general-purpose LLMs translate cultural concepts from Latin, using the epigrams of Martial as a case study. To quantify the nature of these translation shifts, manual annotation was conducted according to the Multidimensional Quality Metrics (MQM) framework, facilitated by adapted categories and a custom-developed web application. In this process, translations generated by the models Gemini 3.1 Pro, M365 Copilot Chat, and Kimi-K2.6 were compared against the source text and two human reference translations. The analysis reveals, on the one hand, a clear tendency towards more literal, generic translations that lack cultural nuance, and, on the other hand, an assimilative translation strategy, presumably driven by an English-language bias in the training data. In the model comparison, Gemini 3.1 Pro produced significantly fewer shifts than the other two models. Finally, the results demonstrate that longer source texts result in a proportionally higher number of shifts per verse.

Inhoudsopgave

Lijst van tabellen	1
Lijst van figuren	1
1) Inleiding	2
2) Literatuurstudie	4
2.1) De opkomst van GenAI en Large Language Models	4
2.2) Overzicht van relevante LLM's	6
2.3) Meertaligheid en bias in LLM systemen	9
2.4) LLM's en Latijn	12
2.5) Een vertaalwetenschappelijk perspectief	14
3) Methode	17
3.1) Multidimensional Quality Metrics	17
3.2) Martialis' <i>Epigrammata</i>	21
3.3) Keuze van professionele vertalingen	23
3.4) Uitvoering	24
4) Analyse	26
4.1) Belangrijkste verschuivingen in de LLM-vertalingen	27
4.2) Modelvergelijking	39
4.3) Epigramlengte	40
5) Discussie	42
6) Conclusie	46
Literatuurlijst:	47

Lijst van tabellen

Tabel 1: Overzicht van de annotaties per categorie.	27
Tabel 2: Overzicht van overly literal shifts.	27
Tabel 3: Voorbeeld bij overly literal shifts (1).	27
Tabel 4: Voorbeeld bij overly literal shifts (2).	28
Tabel 5: Voorbeeld bij overly literal shifts (3).	28
Tabel 6: Voorbeeld bij overly literal shifts (4).	29
Tabel 7: Voorbeeld bij overly literal shifts (5).	29
Tabel 8: Overzicht van undertranslation shifts.	30
Tabel 9: Voorbeeld bij undertranslation shifts (1).	30
Tabel 10: Voorbeeld bij undertranslation shifts (2).	31
Tabel 11: Overzicht van overtranslation shifts.	31
Tabel 12: Voorbeeld bij overtranslation shifts (1).	32
Tabel 13: Vergelijking van de annotatiecategorieën overtranslation en undertranslation.	32
Tabel 14: Overzicht van addition shifts.	33
Tabel 15: Voorbeeld bij addition shifts (1).	33
Tabel 16: Voorbeeld bij addition shifts (2).	33
Tabel 17: Overzicht van omission shifts.	34
Tabel 18: Vergelijking van de annotatiecategorieën addition en omission.	34
Tabel 19: Overzicht van mistranslation shifts.	35
Tabel 20: Voorbeeld bij mistranslation shifts (1).	35
Tabel 21: Voorbeeld bij mistranslation shifts (2).	35
Tabel 22: Overzicht van culture-specific reference shifts.	36
Tabel 23: Voorbeeld bij culture-specific reference shifts (1).	36
Tabel 24: Voorbeeld bij culture-specific reference shifts (2).	37
Tabel 25: Voorbeeld bij Andere verschuivingen (1).	37
Tabel 26: Voorbeelden bij Andere verschuivingen (2).	38
Tabel 27: Voorbeeld bij onbehandelde verschuivingen.	39
Tabel 28: Belangrijke statistieken voor de vergelijking van de LLM's Gemini 3.1 pro, M365 Copilot Chat en Kimi-K2.6.	40

Lijst van figuren

Figuur 1: Gemiddeld aantal punten per vers in verhouding tot de lengte van het epigram... 41

1) Inleiding

Generatieve Artificialle intelligentie (GenAI) en Large Language Models worden steeds vaker gebruikt als taalinstrument in diverse maatschappelijke en academische domeinen. Hierdoor stijgt het belang om meer kennis te verwerven over de werking, het potentieel en de inherente beperkingen ervan. Hoewel het academisch discours in toenemende mate aandacht besteedt aan de risico's van Large Language Models (LLM's), zijn de precieze implicaties ervan nog onvoldoende in kaart gebracht. Zo resulteren de marginalisatie van *low-resource*-talen en de dominantie van een Engelstalige bias in een ongelijke representatie van culturen en hun gedachtegoed, wat kan leiden tot ethnocentrisme (Bender et al., 2021; Navigli et al., 2023).

Bij klassieke talen zoals het Oudgrieks en het Latijn tekenen deze effecten zich scherp af door de grote culturele kloof en de beperkte omvang van de beschikbare trainingscorpora. Toch is er op dit vlak nog nauwelijks onderzoek verricht binnen de klassieke filologie (Strachan, 2023; Volk et al., 2024). Dit onderzoek gebruikt daarom het Latijn als casus voor een *low-resource*-taal, om inzicht te bieden in de impact van deze verhoudingen op machinale vertalingen. De onderzoeksvraag luidt dan ook: 'Op welke manier wordt cultureel ideeëngoed in de Latijnse epigrammen van Martialis vertaald door LLM's?' Door verschuivingen in LLM-vertalingen gedetailleerd te analyseren, met een specifieke focus op de weergave van cultuurspecifieke concepten, biedt deze studie een manier om inzicht te verwerven in tendensen die relevant zijn voor wetenschappelijk onderzoek naar LLM's en vertalingen uit *low-resource*-talen.

De scriptie is als volgt opgebouwd: na het schetsen van het theoretisch kader, waarin de fundamentele concepten en literatuur worden besproken, volgt een uiteenzetting van de gehanteerde methode. De kern hiervan bestaat uit een handmatige annotatie in de ontwikkelde onderzoeksapplicatie op basis van het Multidimensional Quality Metrics-framework (MQM Foundation, 2026). Deze methode wordt toegepast op een corpus van Martialis' epigrammen (Martial, 1993), geselecteerd vanwege de hoge dichtheid aan historische en culturele referenties, die een gekende uitdaging vormen voor LLM's (Bender et al., 2021; Zhong et al., 2026).

Tijdens dit evaluatieproces worden de LLM-vertalingen vergeleken met de Latijnse brontekst en twee menselijke referentieveralingen om tekstuele verschuivingen te identificeren en te classificeren. In de daaropvolgende analysefase worden de resultaten gepresenteerd volgens de belangrijkste annotatiecategorieën. Ten slotte plaatst de discussie deze bevindingen in een breder perspectief, worden de beperkingen van het onderzoek besproken en worden concrete suggesties voor vervolgonderzoek geformuleerd.

Daarmee beoogt deze studie een gefundeerde bijdrage te leveren aan het debat over de culturele sensitiviteit van LLM's en hun omgang met *low-resource*-talen zoals het Latijn.

2) Literatuurstudie

Dit hoofdstuk vormt de theoretische fundering van het onderzoek. Allereerst wordt de technologische evolutie van GenAI en de onderliggende werking van LLM's beknopt uiteengezet. Deze achtergrondkennis is essentieel om de problematieken rond machinale vertalingen, die later centraal staan, te kunnen doorgronden. Dan volgt een overzicht van relevante LLM's waarin de modelselectie wordt verantwoord. Vervolgens worden de fenomenen LLM-bias en de uitdagingen rond *low-resource*-talen binnen de bestaande academische literatuur gesitueerd, waarna de verhouding tussen LLM's en de Latijnse taal wordt belicht. Tot slot biedt dit hoofdstuk een vertaalwetenschappelijk perspectief op betekenisverschuivingen en etnocentrische of domesticerende vertaalstrategieën, geconcretiseerd door voorbeelden van verschuivingen in menselijke vertalingen uit het Latijn.

2.1) De opkomst van GenAI en Large Language Models

Hoewel de snelle evolutie van GenAI pas recentelijk de volle aandacht krijgt, kent de technologie een veel langere ontstaangeschiedenis. Waar de eerste AI-modellen nog steunden op expliciete, handmatige programmering, zorgde de introductie van *machine learning* in de jaren 1950 ervoor dat systemen op basis van data autonoom taken konden uitvoeren. Deze evolutie versnelde aanzienlijk met de komst van *deep learning*, een methode gericht op de verwerking van complexe data via neurale netwerken. Dit zijn computationele modellen die, naar analogie met het menselijk brein, zijn opgebouwd uit meerdere lagen artificiële neuronen (Banh & Strobel, 2023; Noor & Ige, 2025). De implementatie van deze netwerken bleek bijzonder waardevol voor *Natural Language Processing* (NLP), de *machine learning*-discipline die menselijke taal omzet in computerleesbare, gestructureerde data (Galanis et al., 2021). De toepassing van *deep learning* binnen NLP resulteerde in *word embedding*, een techniek waarbij woorden worden omgezet in vectoren die hun positionele en contextuele waarde weerspiegelen (Almeida & Xexéo, 2023). Aan de hand van deze via NLP gegenereerde vectoren zijn generatieve modellen zoals *Recurrent Neural Networks* (RNN's) in staat om telkens het volgende woord in een tekstuele output te berekenen (R. Schmidt, 2019).

Het landschap van GenAI veranderde in 2017, toen Google in het artikel *Attention Is All You Need* de Transformer-architectuur introduceerde, oorspronkelijk bedoeld als opvolger van RNN's (Vaswani et al., 2017). Dankzij *attention mechanisms* kunnen taalmodellen met deze architectuur – ten opzichte van RNN's – focussen op de relevante inputdata, wat de kwaliteit van de output aanzienlijk verbetert (Celik et al., 2025). Sinds de innovatie van Transformers verschoof GenAI van een taakspecifieke benadering naar een algemeen-inzetbare aanpak via *pre-training*. Dat gebeurt aan de

hand van grote, ongelabelde datasets waarmee het model zichzelf traint. Vervolgens kan het model met gelabelde data worden geoptimaliseerd (*fine-tuned*) voor bepaalde taken zonder telkens een nieuw model te hoeven trainen (L. Li et al., 2021; Y. Zhang et al., 2025). De Transformer-architectuur verhoogt aanzienlijk de prestaties van taalmodellen in NLP-opdrachten en vormt de basis voor hedendaagse Large Language Models (Vaswani et al., 2017; Yenduri et al., 2024). In 2018 werd door OpenAI de eerste versie van GPT gelanceerd, een Generatief *Pre-trained* Transformermodel dat uitblonk in het verwerken en genereren van natuurlijk taalgebruik (Radford & Narasimhan, 2018). De daaropvolgende jaren zijn GenAI-modellen steeds verder uitgebreid qua parameters en de grootte van de datasets (Yenduri et al., 2024). Het gevolg van deze schaalvergroting zijn vaardigheden zoals contextueel leren, redeneren en een algemene verbetering in de output van de modellen. De term ‘Large Language Models’ (LLM’s) wordt gebruikt om naar dergelijke modellen te verwijzen vanwege hun grote hoeveelheid parameters en trainingsdata (Brown et al., 2020; Y. Zhang et al., 2025). Zoals in het hoofdstuk ‘Meertaligheid en bias in LLM systemen’ aan bod zal komen, brengt de omvang van LLM-datasets echter ook problemen met zich mee.

Recente ontwikkelingen in GenAI omvatten de vooruitgang in *Retrieval-Augmented Generation* (RAG), technologie die feitelijke incorrectheid van LLM’s vermindert door externe kennis uit documenten te raadplegen op basis van semantische gelijkheid (Gao et al., 2024), alsook multi-modaliteit – het ontvangen, verwerken en genereren van data in vormen zoals afbeeldingen, audio, video en code – (Yin et al., 2024) en *reasoning*-LLM’s. De laatstgenoemde term werd gepopulariseerd bij de lancering van OpenAI o1, een model dat gebruik maakt van versterkend leren en in verschillende stappen redeneert voordat het antwoordt (OpenAI, 2024). Hierdoor kan het model complexe problemen accurater oplossen en verandert de aangewezen *prompting*-strategie van uitgebreide instructies zoals *chain-of-thought prompting* (CoT) naar kortere, duidelijke prompts (Lee et al., 2025; X. Li et al., 2025). Een gedetailleerder overzicht van Large Language Models is te lezen in het artikel *A Comprehensive Overview of Large Language Models*, waarin ook de noodzakelijke onderdelen en de werking van LLM’s uitgebreid worden besproken (Naveed et al., 2024). Daarnaast worden fundamentele concepten uit het LLM-framework bondig uitgelegd in *Comparison of Large Language Models (LLM)* (Celik et al., 2025).

2.2) Overzicht van relevante LLM's

Het veld van LLM's is voortdurend in beweging, waardoor een vergelijkend overzicht van *state-of-the-art* modellen snel gedateerd is. Desondanks heeft elk model een specifieke architectuur en voordelen. Een grondig inzicht in deze verschillen is cruciaal om een geïnformeerde modelselectie te kunnen maken. Om deze twee overwegingen in balans te houden, volgt in dit hoofdstuk een beknopt overzicht van relevante *general-purpose* LLM's, zowel voor deze studie als voor toekomstig onderzoek. De relevantie van deze modellen is geëvalueerd op basis van hun prestaties, hun onderzoeksgerichtheid en de omgang met talige diversiteit en bias. Een populaire methode om prestaties van LLM's met elkaar te vergelijken en vooruitgang vast te stellen, is het gebruik van benchmarks (Reuel et al., 2024) zoals MMLU-PRO (Wang et al., 2024) en Humanity's Last Exam (Phan et al., 2026). Benchmarks zijn gestandaardiseerde testen die de capaciteiten van LLM's (in het algemeen of binnen een specifiek domein zoals 'Wiskunde') meten ten opzichte van een verwachte uitkomst en vervolgens omzetten naar een score. Meestal neemt dit de vorm aan van een *accuracy score* uitgedrukt in een percentage. Hoewel er beperkingen zijn aan bestaande LLM-benchmarks, zoals besproken wordt door Qian et al. (2026), scheppen ze een snel en overzichtelijk beeld van het huidige LLM-landschap.

- ChatGPT door OpenAI staat al sinds de lancering van hun eerste openbaar toegankelijke chatbot in 2022 (Naveed et al., 2024) bekend als een belangrijke speler en pionier op het vlak van mensachtige tekstgeneratie (Celik et al., 2025). De modellen hebben ook multimodale capaciteiten en een breed scala aan toepassingen (Liu et al., 2023). De ChatGPT-reeks wordt vaak vermeld als een vertegenwoordiger van LLM's in brede zin door zowel populaire media als onderzoekers.
- Gemini door Google AI is de opvolger van Bard (Naveed et al., 2024) en was aanvankelijk vooral sterk in multimodale opdrachten (Ahmed & Islam, 2024). Recente modellen hebben echter grote vooruitgang geboekt ten opzichte van de concurrentie op andere functiedomeinen, zoals de AI-benchmarks suggereren in *A new era of intelligence with Gemini 3* (Hassabis et al., 2025). De Gemini-reeks is ook geïntegreerd in het Google-ecosysteem, waardoor het rechtstreeks gebruikt kan worden in onder andere Gmail en Google Docs, (Ahmed & Islam, 2024; Celik et al., 2025).
- Copilot door Microsoft wordt vaak ingezet als hulpmiddel bij administratieve of technische taken van kenniswerkers wegens gebruiksvriendelijkheid en integratie in het Microsoft 365-ecosysteem (Word, Outlook, Teams, etc.) (Dillon et al., 2025; C. F. Schmidt et al., 2026). Bovendien biedt Copilot verhoogde databeveiliging voor bedrijven en instellingen (Lund-Tonnesen et al., 2025).

- Claude door Anthropic is verfijnd door de techniek 'Constitutionele AI', wat inhoudt dat het in de trainingsfase behalve menselijke feedback ook AI-gegenereerde feedback krijgt, gestuurd door de principes van de *constitution*. Door dit proces zouden de modellen minder schadelijke content genereren, wat in lijn ligt met de veiligheids- en ethische overwegingen die Anthropic voorop stelt (Bai et al., 2022; Priyanshu et al., 2024).
- Llama door Meta AI heeft een onderzoeksgericht ontwerp en wordt voornamelijk gekozen voor de goede instructie-tuning en parameterefficiëntie (Celik et al., 2025; Naveed et al., 2024). Dat laatste houdt in dat niet alle parameters worden betrokken voor het finetunen, wat de benodigde rekenkracht vermindert (L. Xu et al., 2023).
- BLOOM door Bigscience is getraind op het ROOTS corpus, dat 46 natuurlijke talen bevat, exclusief sommige *high-resource* talen zoals Koreaans en Russisch. De ontwikkelaars hebben immers ingezet op menselijke betrokkenheid en inclusiviteit bij het trainen, waarbij onder meer data van elke taal betrokken in *pre-training* door experts moesten worden gevalideerd (Yong et al., 2023). De bedoeling van Bigscience is om bias tegenover gemarginaliseerde groepen te beperken in hun LLM (Bigscience Workshop et al., 2023).
- Kimi door Moonshot AI is een reeks van *mixture of experts* (MoE) LLM's gericht op complex redeneren en *Agentic Intelligence* (de vaardigheid om zelfstandig te functioneren in een dynamische omgeving door actief te leren en zich na pre-training nog aan te passen). Een MoE-model is opgedeeld in meerdere 'experten' die ingeschakeld worden naargelang de instructie, wat de snelheid en efficiëntie verhoogt ten opzichte van de berekeningskosten (D. Zhang et al., 2025). Recente Kimi modellen staan momenteel aan de top van *open-source*-LLM's op basis van verschillende benchmarks (Kimi Team et al., 2026).

Alle genoemde modellen zijn gebouwd op de Transformer-architectuur en zijn *decoder-only*, wat betekent dat ze in de eerste plaats ontworpen zijn om tekst te genereren door telkens het volgende token (vergelijkbaar met een woord of woorddeel) te voorspellen (Naveed et al., 2024). Er zijn ook *encoder-only* taalmodellen, zoals BERT door Google AI, die niet behoren tot GenAI, maar wel een grote rol hebben gespeeld in het gerelateerde onderzoeksdomein van Natural Language Processing. Deze modellen zijn niet gericht op tekst produceren, maar eerder op nuttige informatie extraheren en vervolgens analyseren (Lin et al., 2021). Tot slot propageert het onderzoek door B. Zhang et al. (2025) het gebruik van *encoder-*

decoder modellen, een type taalmodel met verrassende prestaties in verscheidene domeinen na uitvoerige finetuning, maar overschaduwd door de populariteit van het *decoder-only* model GPT en zijn prestaties in *zero-shot*- en *few-shot*-contexten.

ChatGPT, Gemini, Copilot en Claude zijn *closed-source*-modellen, waardoor de gebruikte trainingsdata en optimalisatietechnieken slechts gedeeltelijk of helemaal niet openbaar beschikbaar zijn. De meeste *closed-source*-modellen zijn toegankelijk via *application programming interfaces* (API's). Door grootschalige investeringen van de eigenaars presteren deze modellen historisch gezien beter op algemene benchmarks zoals HumanEval, GSM8K, MMLU, DROP, en GPQA-Diamond (Manchanda et al., 2025). Llama, BLOOM en Kimi daarentegen zijn voorbeelden van *open-source*-modellen. Dit betekent dat de architectuur, trainingsdata en algoritmen publiek bekend en beschikbaar zijn. *Open-source*-modellen zijn over het algemeen beter valideerbaar en kunnen door andere onderzoekers verder worden ontwikkeld (Manchanda et al., 2025). Het gebruiksgemak en de prestaties van *closed-source*-modellen enerzijds en de onderzoeksgerichtheid en transparantie van *open-source*-modellen anderzijds is een substantiële afweging voor modelkeuze.

Voor deze studie is een doelbewuste selectie gemaakt van drie geavanceerde Large Language Models: Gemini 3.1 Pro, M365 Copilot Chat en Kimi-K2.6 (*Google Gemini*, z.d.; *Kimi AI*, z.d.; *M365 Copilot*, z.d.). Deze combinatie bevat respectievelijk twee *closed-source*-modellen en een *open-source*-model. Ten tijde van dit onderzoek behoort Gemini 3.1 Pro tot de absolute top volgens toonaangevende benchmarks. Het model behaalde onder meer topscores in *Humanity's Last Exam* (Center for AI Safety et al., 2026) en de *MMLU-Pro* (*Gemini 3.1 Pro Preview - Intelligence, Performance & Price Analysis*, z.d.), een benchmark die gericht is op meertalig redeneervermogen (Wang et al., 2024). De toegang tot deze versie van Gemini is vergrendeld achter een betaald abonnement. De keuze voor M365 Copilot Chat komt voort uit academische overwegingen, aangezien het model databescherming garandeert voor aangesloten universiteiten, waaronder de KU Leuven (*Microsoft Copilot*, 2026). Tot slot is Kimi-K2.6 aan de methodologie toegevoegd als *open-source* alternatief. Dit model behoort tot de best presterende opties in zijn klasse volgens onder meer de *GPQA Diamond* benchmark en *Humanity's Last Exam* (*Kimi K2.6 - Intelligence, Performance & Price Analysis*, z.d.; Vellum, 2026).

2.3) Meertaligheid en bias in LLM systemen

Het gebruik van LLM's gaat gepaard met allerlei risico's en problemen, waaronder hallucinatie, ambiguïteit, onvolledigheid en bias (H. J. Kim et al., 2024; Y. Zhang et al., 2025). Hallucinatie houdt in dat de LLM een plausibele output genereert die afwijkt van de input, de voorafgaande outputs of feitelijke informatie (Y. Zhang et al., 2025). Soms herkent het model dat het in de fout is gegaan, maar beslist het alsnog om verder in te zetten op de foutieve output om consistent te blijven. Dit fenomeen wordt *hallucination snowballing* genoemd (M. Zhang et al., 2023; Y. Zhang et al., 2025). Ambiguïteit bij LLM's slaat op het genereren van informatie die niet noodzakelijk incorrect is, maar wel voor meerdere interpretaties vatbaar is. Een onvolledige of gefragmenteerde output is ook mogelijk, meestal veroorzaakt door onvolledigheid in de dataset zelf of door een te lange input (Y. Zhang et al., 2025). De term bias binnen de context van dit onderzoek verwijst naar 'the fact of a collection of data containing more information that supports a particular opinion than you would expect to find if the collection had been made by chance' (Cambridge Dictionary, 2026).

Wanneer een LLM dus een output genereert die gekleurd is door overheersende opvattingen uit een bepaalde cultuur, gemeenschap of ideologie, is deze output onderhevig aan bias (Bender et al., 2021; Y. Zhang et al., 2025). Zo kan een biased output bijvoorbeeld stellen dat beroep X samenhoort met bevolkingsgroep Y of een ongepaste vertaling maken van een cultuurgebonden begrip. Resnik (2025) vestigt de aandacht op twee elementen: enerzijds keuze – de LLM kiest een bepaalde formulering of vertaling – en anderzijds de invloed van voorafgaande informatie op die keuze (p. 2). Binnen het kader van LLM-vertalingen is een biased output geheel onwenselijk, ongeacht de gebruikscontext, omwille van meerdere negatieve gevolgen. Hoe kleiner de (digitale) aanwezigheid van een taal en hoe groter de culturele kloof ten opzichte van dominante talen, des te belangrijker is het om de impact van bias op die taal te beperken. Voorlopig zijn zulke *low-resource*-talen, waartoe ook het Latijn behoort (Hudspeth et al., 2026), echter het primaire object van LLM-bias (Navigli et al., 2023; Y. Xu et al., 2025).

Zoals eerder vermeld, gebruiken LLM's een enorme hoeveelheid trainingsdata om hun prestaties te verbeteren. De toegankelijkste en rijkste bron voor die data is het internet, maar dit creëert geenszins een perfecte dataset. Toegang tot het internet is immers gelimiteerd door allerlei factoren, zoals een gebrek aan (voldoende) draadloos netwerk, de prijs van technologie en onvoldoende geletterdheid (Yu et al., 2024). Hierdoor zijn bepaalde groepen van mensen slechts deels of niet gerepresenteerd in de data. Hoewel het praktisch onmogelijk is om alle internetgebruikers te identificeren, zijn er indicaties dat zij die wél internettoegang hebben evenmin de demografische posities gelijkmatig vertegenwoordigen (Navigli et al., 2023). Een enquête door Pew Internet Research in 2016 (Barthel et al., 2016) toont bijvoorbeeld dat de gebruikers

van Reddit, een website die een grote invloed had op de trainingsdata van GPT-2, voornamelijk mannen waren van onder de 30 jaar (67% van de gebruikers in de Verenigde Staten). De ongelijke verdeling van internetgebruikers leidt tot een oververtegenwoordiging van dominante opvattingen die bias veroorzaken in LLM's (Navigli et al., 2023). Dat is nadelig voor gemarginaliseerde bevolkingsgroepen en hun ideeëngoed (Bender et al., 2021) en kan zelfs resulteren in racisme. Hoffman et al. (2024) maken aannemelijk dat Large Language Models vanwege dialectvooroordeel andere (hypothetische) beslissingen maken over de sprekers van bepaalde dialecten. Zo werden door de LLM's, in theoretische scenario's, sprekers van African American English gemiddeld een minder prestigieuze job toebedeeld en vaker met de doodstraf bestraft dan sprekers van Standardized American English (Hoffman et al., 2024).

De bias manifesteert zich niet enkel op het gebied van ideeën en opvattingen, maar heeft ook een impact op talen zelf. De trainingsdata voor LLM's zijn namelijk grotendeels in het Engels (Bender et al., 2021) en dit zorgt voor grote verschillen in performance afhankelijk van de taal waarmee een LLM aan de slag gaat. (Kocmi et al., 2023; Rigouts Terryn & de Lhoneux, 2024). Voor GPT-3 (2020) waren zo'n 93% van de trainingsdata in het Engels op basis van woordenaantal (Brown et al., 2020). Bovendien zouden LLM's intern redeneren in een taal die sterk op het Engels lijkt vooraleer ze een output genereren in een niet-Engelse taal (Rigouts Terryn & de Lhoneux, 2024; Wendler et al., 2024). In het onderzoeksartikel *Do Multilingual LLMs Think In English?* (Schut et al., 2025) stellen de auteurs evenzeer dat 'multilingual LLMs perform key reasoning steps in a representation that is heavily shaped by English in a way that is not transparent to system users.' (Rigouts Terryn & de Lhoneux, 2024). De dominantie van de Engelse taal in het ecosysteem van LLM's vormt een potentieel gevaar voor talen met minder beschikbare data (*low-resource languages*) en het gedachtengoed dat ermee gepaard gaat. Deze mening wordt ondersteund door Navigli et al. (2023), die schrijven dat '... the gap in quantity, quality, and also diversity (e.g., of annotations) between the text available in high-resource languages and *low-resource* languages is becoming increasingly wider' (p. 6). Door het toenemend gebruik van Generatieve AI is er steeds minder ruimte voor *low-resource*-talen als communicatiemiddel en als bewaarplaatsen van culturele kennis, historische narratieven en wereldvisies (Bender et al., 2021; Zhong et al., 2026).

Large Language Models kunnen echter ook ingezet worden om deze evolutie tegen te gaan, namelijk als virtuele assistenten die helpen om *low-resource*-talen te vertalen en te analyseren. Daardoor kunnen meer onderzoekers en taalliefhebbers met deze talen aan de slag, wat op zijn beurt voordelig is voor de studie en het behoud ervan (Zhong et al., 2026). Voorlopig zijn LLM's echter nog beperkt in vaardigheden die cruciaal zijn om dit te verwezenlijken, zoals het vertalen van metaforen, humor, beleefdheidsvormen en andere taalkundige fenomenen die diep geworteld zijn in een

culturele context. Dit vereist niet enkel oppervlakkige taalkennis, maar ook inzicht in de onderliggende context die mee de betekenis construeert. Engelse bias hindert LLM's om cultuurgebonden uitdrukkingen hun ware betekenis toe te kennen en te begrijpen, met misverstanden en foute interpretaties als gevolg (Navigli et al., 2023; Zhong et al., 2026).

Een aantal methoden om de prestaties van LLM's te verbeteren bij *low-resource*-talen worden vermeld door Zhong et al. (2026), waaronder *transfer learning*, het overbrengen van taalkundige informatie uit trainingsdata van *high-resource* talen naar *low-resource*-talen, en *cross-language pre-training*, het trainen op meertalige data om gedeelde representaties tussen talen te leren. Deze methoden hebben echter betrekking op *pre-training* van nieuwe LLM's en zijn moeilijk inzetbaar bij gebruik van bestaande LLM's. Daarnaast lijkt het finetunen van een LLM voor een specifieke taal bias te verminderen (Rigouts Terryn & de Lhoneux, 2024), alsook het gebruiken van synthetische data in *low-resource*-talen bij *pre-training* (Nadas et al., 2025). Synthetische data zijn kunstmatig gegenereerd door algoritmen of generatieve modellen om werkelijke data na te bootsen (Nadas et al., 2025).

Bender et al. (2021) en Resnik (2025) beargumenteren echter dat bias in LLM's niet volledig kan verdwijnen met de huidige aanpak, omdat de fundamentele architectuur van Large Language Models niet toelaat een grondig onderscheid te maken tussen 'neutrale' co-existentie van concepten (bijvoorbeeld 'Soldaat' [x] en 'Leger' [y]) en cultuurgebonden contingente feiten (bijvoorbeeld 'Soldaat' [x] en 'Man' [y]). In beide scenario's treden begrippen x en y vaak samen op, maar de tweede variant is een vorm van genderbias. Het artikel *Should We Ban English NLP for a Year?* adviseert om de Engelse dominantie in LLM's op een andere manier te bestrijden, namelijk met juridisch-economische maatregelen tegenover ontwikkelaars en gebruikers van Engelse LLM's (Søgaard, 2022). De urgentie om de huidige situatie te verbeteren overstijgt immers het risico dat LLM-bias alsnog zal blijven bestaan.

Om bias te identificeren, wordt dikwijls gebruik gemaakt van *Template-based probes* (Huang et al., 2020; Liang et al., 2021), waarin voorbeeldzinnen staan die door de LLM moeten worden aangevuld met attributen al dan niet gerelateerd aan etniciteit, gender, of een ander sensitief domein (bijvoorbeeld '*I'm sick of the {race_adj} stereotypes*'). In het artikel *Template-Based Probes Are Imperfect Lenses for Counterfactual Bias Evaluation in LLMs* wordt echter geopperd dat deze methode een vals idee geeft van bias vanwege taalkundige gemarkeerdheid in de trainingsdata (Kohankhaki et al., 2026). Taalkundige gemarkeerdheid is een morfologisch fenomeen dat optreedt bij het verwijzen naar een bepaalde groep ten opzichte van de gepercipieerde standaardgroep, zoals 'actrice' ten opzichte van 'acteur' of 'zwarte president' ten opzichte van 'president'. Bias in LLM's wordt met *Template-based*

probes minder opgemerkt omdat de ‘afwijkende’ groepen, omwille van hun gemarkeerdheid, vaker gerepresenteerd worden in de templates dan in een normale setting (Kohankhaki et al., 2026). Karpouzis (2024) presenteert LLM-bias eveneens als een probleem van gelimiteerde en verstoorde representatie en bespreekt enkele methoden voor het vaststellen ervan, zoals sentiment analysis, een vorm van corpusanalyse waarin positieve of negatieve sentimenten worden onderzocht in verhouding tot bevolkingsgroepen. Tot slot bieden Rigouts Terryn & de Lhoneux (2024) een mogelijke annotatiestructuur voor de gevolgen van Engelse bias bij LLM-vertaling in het Nederlands en Frans. Het identificeren en analyseren van de gevolgen van bias is een noodzakelijke stap om de negatieve impact ervan op *low-resource*-talen zoals Latijn te verminderen.

2.4) LLM's en Latijn

Klassieke talen, waaronder het Latijn, zijn een uniek obstakel voor LLM's door hun gelimiteerde corpora, archaïsche grammatica en onvolledige digitalisatie (Zhong et al., 2026). Talen met naamvalsmarkering en vrije woordvolgorde zijn bovendien moeilijker te vertalen voor LLM's (Rosenthal, 2023). Daarnaast is er voor antieke teksten begrip nodig van de historische context, die dikwijls ver verwijderd is van de hedendaagse culturele context. Het correct begrijpen en weergeven van historische allusies, symbolisme en elementen uit de orale traditie zijn extra uitdagingen bovenop de eerder genoemde problemen bij *low-resource*-talen (Zhong et al., 2026). Yan et al. (2024) merken verder op dat de vertaalprestaties van GPT-4 lager zijn bij een *low-resource*-talenpaar dan bij combinaties met Engels. In het artikel *Better To Ask in English? Evaluating Factual Accuracy of Multilingual LLMs in English and Low-resource Languages* worden vergelijkbare bevindingen gepresenteerd voor de Indische talen ten opzichte van het Engels (Rohera et al., 2025).

Gericht onderzoek naar bias voor vertalingen van Latijn door LLM's is nog beperkt, maar over het algemeen wordt verondersteld dat ook het Latijn object is van bias door de dominante representatie van *high-resource* talen in Large Language Models (Strachan, 2023; Zhong et al., 2026). Zoals besproken in de vorige alinea, zijn het gebrek aan historische context en culturele kennis eveneens een aanzienlijke storing voor LLM's. Dat maakt Latijnse vertalingen vatbaar voor biased interpretaties die eventuele lacunes in de domein-specifieke kennis van het model opvullen (Zhong et al., 2026). Ondanks deze hindernissen lijken LLM's vergelijkbaar of zelfs beter te presteren dan concurrerende alternatieven bij het vertalen en interpreteren van klassieke talen. Zo wordt in het artikel *LLM-based Machine Translation and Summarization for Latin* (Volk et al., 2024) het model GPT-4 vergeleken met andere neurale Machine Translation (MT) technologie bij het vertalen van zestiende-eeuwse brieven in Latijn en Vroegnieuwhoogduits naar modern Engels en Duits. De resultaten

van deze vergelijking wijzen op de superioriteit van GPT-4 over de andere geteste MT-systemen, waaronder Google Translate. Het gaat dan vooral over vertaalkwaliteit, robuustheid in meertalige contexten en de mogelijkheid om de vertaling bij te sturen met aanvullende instructies (Volk et al., 2024).

De keuze tussen *general-purpose* LLM's zonder finetuning of modellen gespecialiseerd in het vertalen van Latijn is voorlopig een moeilijke kwestie. Indien de data buiten het trainingsdomein van het gespecialiseerde model liggen, levert een *general-purpose* LLM vaak gelijkwaardige of betere resultaten voor klassieke talen (Akavarapu et al., 2025). Een studie naar de vertaalcapaciteit voor Latijn van ChatGPT 3.5 (november 2022) stelt dat het model een acceptabele output genereert voor een beginnende student Klassiek Latijn (Ross, 2023). Volgens een berekende schatting door Burns (2023) is dit model getraind op zo'n 339 miljoen tokens in het Latijn, inclusief tokens uit Neolatijn, Vicipaedia, en andere niet-klassieke bronnen (G. Schmidt et al., 2024). Onderzoek naar de vertaalcapaciteiten van recentere modellen of van andere LLM's dan de ChatGPT-reeks voor Latijn is momenteel nog schaars (Volk et al., 2024).

Anderzijds hebben ook gespecialiseerde alternatieven bijzondere prestaties geleverd in Latijnse Machine Translation. Hoewel er binnen het kader van dit onderzoek is gekozen voor *general-purpose* LLM's om de reikwijdte en relevantie van de resultaten te verbreden, bestaan er andere opties die in vervolgstudies verder geëxploreerd kunnen worden. In 2020 werd het encoder-only Transformer model LatinBERT uitgebracht, getraind op 642.7 miljoen woorden en bepaalde dat een nieuwe *state of the art* voor Latijnse part-of-speech tagging en het voorspellen van ontbrekende tekst (Bamman & Burns, 2020; Volk et al., 2024). De pipeline ontwikkeld in het onderzoek "*Don't Teach Minerva*": *Guiding LLMs Through Complex Syntax for Faithful Latin Translation with RAG* maakt gebruik van een gespecialiseerd MT model (NLLB) om een eerste draft te generen van de vertaling, die vervolgens door een *general-purpose, open-source*-LLM (Llama-3.3 of Qwen3) wordt verfijnd. Deze methode voor Latijnse vertaling presteert vergelijkbaar met de GPT-5 baseline (Aguilar, 2025) en is interessant voor toekomstig onderzoek rond Latijnse MT vanwege het *open-source* karakter. Ten slotte blijkt dat het combineren van meerdere LLM's in een gemeenschappelijk framework de prestaties kan verhogen voor complexe taken, inclusief het vertalen van Latijn, zoals geïllustreerd in het project LITTERA (*Latin Interpretation and Translations into English for Research Assistance*) (Rosu, 2025).

2.5) Een vertaalwetenschappelijk perspectief

Reflecties over vertalingen kwamen reeds voor in de Oudheid en waren voornamelijk gekaderd binnen het debat tussen ‘letterlijke’ en ‘vrije’ vertaling dat nog eeuwen gevoerd zou worden. Invloedrijke Romeinse auteurs zoals Quintus Horatius Flaccus (65-8 v.C.) en Marcus Tullius Cicero (106-43 v.C.) waren tegenstanders van ‘letterlijke’ vertalingen en pleitten voor een esthetische, creatieve vertaling van de brontekst (Munday, 2016). Amos (1920) beargumenteert echter dat beoefenaars van vroege vertaaltheorie nauwelijks terugkeken naar voorafgaande stellingen en standpunten, waardoor vooruitgang in dit domein zeldzaam was. Bijgevolg bleven concepten en de terminologie om over vertaling te schrijven en oordelen (‘creatief’, ‘accuraat’, ‘trouw’, ‘vrij’, etc.) tot de moderne tijd relatief onnauwkeurig en voor interpretatie vatbaar (Amos, 1920). Dit veranderde in de tweede helft van de twintigste eeuw, toen de academische discipline van vertaalwetenschap tot stand kwam. Stelselmatig begonnen vertaaltheoretici bestaande concepten te herdefiniëren in wetenschappelijke termen en werden vertalingen descriptief beoordeeld in plaats van normatief (Munday, 2016).

Binnen deze studie zijn de theorieën uit ideologische hoek uitermate belangrijk, zoals door Antoine Berman (1942-1991), een Franse vertaler en vertaaltheoreticus. Hij contrasteerde zich met eerdere linguïstische benaderingen zoals door Eugene Nida (1914-2011) door zich af te zetten tegen de notie van dynamische equivalentie – hetzelfde effect proberen tweebrengen bij de lezers in doeltaal als bij de lezers in de brontaal – vanwege het verlies dat dit tweebrengt voor de eigenheid van de brontekst. Het focussen op de boodschap en het verwerpen van de wezenlijke tekst als een soort verpakkingsmateriaal zou immers bijdragen aan de dominantie van ethnocentrisme in de westerse vertaaltraditie (Venuti, 2021). Deze kritiek is vergelijkbaar met de problematiek rond bias en LLM’s vandaag en is daarom relevant binnen het theoretisch kader van dit onderzoek.

Lawrence Venuti bouwde verder op Bermans ideeën in zijn werk *The Translator’s Invisibility* (1995), dat een nog sterkere ideologische inslag heeft. Volgens Venuti wordt de Anglo-Amerikaanse vertaalcultuur gedomineerd door een domesticerende vertaalstrategie, met risico op ethnocentrisme en zelfs racisme als gevolg. Het concept ‘domesticatie’ is een uitbreiding van wat Berman ‘ethnocentrische vertaling’ noemt. Engelse vertalers verwaarlozen onder het mom van ‘fluency’ en ‘readability’ de eigenheid van de brontekst. De remedie is een *foreignizing* vertaalstrategie, die de nadruk legt op de taalkundige en culturele verschillen tussen brontekst en doelttekst (Venuti, 2017). Er is nog weinig tot geen wetenschappelijk onderzoek waarin Venuti’s ideeën worden toegepast op LLM-vertalingen, ondanks de toenemende aandacht voor cultural bias bij Large Language models (Karpouzis, 2024; Navigli et al., 2023; Resnik, 2025). In verband met *domesticating* en *foreignizing* vertaalstrategieën,

stellen Yang & Yang (2026) wel dat ‘LLMs, including state-of-the-art models, often lack training in these strategies, resulting in issues such as mistranslation, cultural mistranslation, and omission.’ (p. 8). Door middel van een grondige analyse, met een focus op de weergave van cultureel gedachtegoed, tracht deze studie licht te werpen op de taalkundige en culturele verschillen tussen brontekst en doelttekst in LLM-vertalingen.

Om die verschillen tussen bron- en doelttekst in LLM-vertalingen uit het Latijn adequaat te kunnen analyseren, is het essentieel om eerst de specifieke uitdagingen van menselijke vertaalpraktijken vanuit deze klassieke taal in kaart te brengen. Een prominent vraagstuk in dit kader is dat van de domesticerende vertalingen. Die kwamen immers reeds voor bij vertalingen door de Romeinen zelf. Zo vertaalde Livius Andronicus de Ὀδύσσεια (*Odýsseia*) door Homeros niet in de oorspronkelijke Griekse dactylische hexameters, maar in inheemse Saturnische verzen, wat ongetwijfeld de culturele ‘vreemdheid’ van het originele werk (deels) moet hebben verhuld en het Romeinse publiek een meer vertrouwde lees- of luisterervaring hebben geboden. (McElduff, 2013)

Een vergelijkbare domesticerende tendens in een modernere context wordt door Venuti (2017) geïllustreerd aan de hand van Robert Graves' Engelse vertaling van Suetonius' *De vita Caesarum*. Graves hanteerde een verregaande assimilatiestrategie om de tekst vlotter en toegankelijker te maken voor het Engelse lezerspubliek. Zoals hij in zijn voorwoord aangeeft, deinsde hij er niet voor terug om de Latijnse zinsstructuren drastisch te herschikken, verklarende elementen direct in de lopende tekst te integreren ter vervanging van voetnoten, en cultuurspecifieke concepten zoals dateringen, toponiemen en munteenheden te moderniseren (Graves, 1957). Venuti wijst echter op het gevaar van dergelijke ingrepen: naast het verlies van de eigenheid van de brontekst, resulteert deze methode vaak in een vertekende weergave door de introductie van anachronistische en etnocentrische waarden. Een voorbeeld hiervan is te vinden in Graves' vertaling van de zinsnede *prostratae regi pudicitiae* (letterlijk: 'hij wierp zijn kuisheid neer voor de koning'), waar hij de term 'homosexual relationship' gebruikt. Deze vertaalkeuze is fundamenteel anachronistisch, aangezien de Romeinse seksualiteitsbeleving niet werd gecategoriseerd op basis van biologisch geslacht, maar op basis van de actieve of passieve rol van de participanten (Venuti, 2017).

Naast deze cultuurhistorische aanpassingen benadrukt André Lefevere (1975) in zijn invloedrijke analyse van literaire vertalingen van Catullus een reeks linguïstische en semantische valkuilen. Een veelvoorkomende strategie is het gebruik van archaïsmen ter compensatie van een ontbrekend hedendaags equivalent, waarmee vertalers de grenzen van hun eigen doeltaal trachten te rekken. Dit brengt echter het risico met

zich mee dat de semantiek van het archaïsme doorheen de tijd is verschoven. Lefevre illustreert dit met het Latijnse *bustum*, een plaats voor lijkverbranding en begraving, wat in de tijd van Catullus geen archaïsme was. Een Engelse vertaling als *pile* verliest de connotatie van dood en begraven, terwijl een ogenschijnlijk passend archaïsme als *barrow* ('grafheuvel') in modern Engels evengoed 'kruiwagen' kan betekenen. Een vergelijkbaar betekenisverlies treedt op wanneer vertalers, uit filologische of louter praktische overwegingen, teruggrijpen naar etymologisch verwante equivalenten die sindsdien een betekenisverschuiving hebben ondergaan, zoals de letterlijke omzetting van *silvestria dona* naar *silvestrian gifts*. Zelfs bij een volledige semantische equivalentie kan de communicatieve waarde van de vertaling nog steeds afwijken van het origineel. De vertaling van het Latijnse *pulvinar* ('goddelijk rustbed') naar het alledaagse Engelse *couch* reduceert bijvoorbeeld het sacrale aspect (Lefevre, 1975).

Bij het vertalen van bronteksten die temporeel, ruimtelijk en traditioneel ver verwijderd zijn van de doelcultuur, zoals Latijnse teksten uit de Oudheid, stuiten vertalers voortdurend op een precaire balans. Zij worden meestal gedwongen om een compromis te zoeken tussen het accuraat overbrengen van de feitelijke betekenis enerzijds en het bewaren van de oorspronkelijke communicatieve waarde anderzijds. In dit proces moet steeds de afweging worden gemaakt in hoeverre de eigenheid van de brontekst behouden blijft, geherinterpreteerd wordt, of veeleer wordt genaturaliseerd door middel van direct begrijpelijke formuleringen in de doeltaal (Lefevre, 1975). Bij het bespreken van de methode zal vervolgens nader worden toegelicht hoe de analyse en beoordeling van deze specifieke vertaalproblematieken wordt uitgevoerd binnen het onderzoek naar de vertalingen van Latijnse teksten door Large Language Models.

3) Methode

3.1) Multidimensional Quality Metrics

Zoals vermeld in de inleiding, ligt de focus van dit onderzoek bij het beschrijven en analyseren van verschuivingen ten opzichte van de Latijnse brontekst, niet bij een error-analysis van LLM-vertalingen (ondanks de formele gelijkenissen). De gehanteerde onderzoeksmethode betreft dan ook een handmatige annotatie binnen het Multidimensional Quality Metrics (MQM)-framework, dat aangepast is aan de scope van deze studie. In de volgende alinea wordt beknopt een populaire alternatieve analysemethode besproken, om vervolgens over te gaan op de motivering en werking van MQM-annotatie.

Die alternatieve methoden om de vertaalprestaties van LLM's te meten, zoals BLEU (bilingual evaluation understudy) en ChrF, zijn automatisch en hebben als voordeel dat ze erg snel resultaten leveren en minder arbeidsintensief zijn, waardoor ze nuttig zijn voor ontwikkelaars van MT-systemen die frequent het effect van hun ingrepen moeten testen (Dabre et al., 2025; Papineni et al., 2002; Popović, 2015; Wang & Yu, 2025). Hoewel BLEU en ChrF erg prominent zijn in onderzoek over vertaalkwaliteit door LLM's (Rosenthal, 2023; Rosu, 2025; Sutskever et al., 2014; Vaswani et al., 2023), hebben ze een aantal nadelen ten opzichte van handmatige annotatie. Ten eerste werken deze automatische methoden volgens een vast algoritme, waardoor ze minder flexibel zijn voor mensen zonder gevorderde programmeervaardigheden. Ten tweede leveren ze slechts één (numerieke) score als resultaat en bieden ze bijgevolg nauwelijks inzicht in de reden waarom een vertaling een bepaalde score heeft gekregen. Ten slotte missen automatische methoden cruciale semantische kennis, waardoor equivalente synoniemen en parafrase soms als 'fouten' worden geannoteerd (Blagec et al., 2022; Fernandes et al., 2023; Y. Li et al., 2024).

Om de betrouwbaarheid van handmatige annotatie te waarborgen, is een gestandaardiseerd framework vereist opdat de onderzoeksresultaten reproduceerbaar en empirisch vergelijkbaar zijn. Het MQM-framework biedt een gestructureerde methodologie voor identificatie en classificatie van vertaalfouten en -verschuivingen en is bovendien aanpasbaar aan het specifieke onderzoek waarbinnen het wordt toegepast (Freitag et al., 2021; Mariana et al., 2015). De mogelijke annotaties zijn per categorie gerangschikt waarbij elke categorie hiërarchisch is ingedeeld. Door variaties individueel te annoteren en hieraan vooraf gedefinieerde gewichten toe te kennen naargelang de intensiteit, worden overhaaste of willekeurige beoordelingen voorkomen (Freitag et al., 2021). MQM is initieel ontworpen voor het evalueren van zowel menselijke vertalingen als van machine translation en kan dus ook worden toegepast op GenAI-vertalingen (Lommel et al.,

2024). Hierdoor heeft het framework recent aan populariteit gewonnen en is het zelfs geïntegreerd in de *Workshop on Machine Translation* (Freitag et al., 2024), waar het dienst doet als benchmark voor automatische evaluatiemethoden (Fernandes et al., 2023).

Voorts is MQM een beproefde methode in wetenschappelijk onderzoek. Zo gebruiken Yan et al. (2024) MQM om menselijke vertalers te vergelijken met GPT-4 voor vier verschillende talen (Engels, Hindi, Russisch en Chinees). Ten slotte gebruiken Yang en Yang (2026) MQM-annotatie om literaire LLM-vertalingen uit het Chinees te evalueren, met aandacht voor culturele elementen. Waar hun onderzoek zich echter beperkt tot één enkele annotatiecategorie voor cultuurgebonden afwijkingen, beoogt de huidige studie een diepgaandere analyse door de hiërarchische structuur en flexibiliteit van het MQM-raamwerk optimaal te benutten. Deze eigenschappen, samen met het toenemend gebruik van MQM binnen de evaluatie van automatische vertalingen, onderbouwen de keuze voor deze methode.

Het MQM-framework kan ook worden ingezet voor automatische zelfevaluatie van Large Language Models, een paradigma dat doorgaans wordt aangeduid als *LLM-as-a-judge* (LAJ). Binnen deze methodiek fungeert een LLM autonoom als evaluator om een model-output te evalueren aan de hand van vooraf gedefinieerde criteria binnen een specifieke evaluatiemetriek (A. Kim, 2025). Het MQM-framework biedt een gestructureerde basis voor het vastleggen van deze evaluatiecriteria (Fernandes et al. 2023; A. Kim 2025). LAJ-evaluatie heeft echter inherente beperkingen zoals hallucinatie, onvoldoende kennis in domein-specifieke situaties, afnemende nauwkeurigheid bij lange inputs, en invloed door bias (Domhan & Zhu, 2025; H. Li et al., 2024), waardoor het niet geschikt is voor deze studie. In het onderstaande gedeelte wordt beschreven hoe een handmatige MQM analyse moet worden uitgevoerd.

De eerste fase omvat de selectie van de te analyseren tekst. Vervolgens moet de MQM-typologie worden aangepast door relevante categorieën en subcategorieën uit de uitgebreide hiërarchie te selecteren. Deze afbakening wordt afgestemd op de specifieke doelstellingen van het onderzoek en de eigenschappen van de tekst, rekening houdend met onder meer het documenttype, het doelpubliek en het gehanteerde register. Tijdens de analyse zelf wordt de tekst systematisch doorzocht op afwijkingen. Elke geïdentificeerde afwijking krijgt een annotatie op basis van de geselecteerde MQM-categorieën en een bijbehorende ernstgraad. Hoewel het standaardraamwerk vier niveaus hanteert (*neutral*, *minor*, *major* en *critical*), wordt het niveau *neutral* in dit specifieke onderzoek enkel sporadisch gebruikt om overbodige tekst die tussen haakjes staat in de vertaling en sommige repeterende verschuivingen aan te duiden. Deze ernstgraad dient normaal gezien om afwijkingen van de

persoonlijke voorkeur van de evaluator te markeren bij vertalingen of teksten. Het vergelijken met gepubliceerde menselijke vertalingen is echter een objectievere manier om dit soort afwijkingen te analyseren, annoteren en bespreken. Daarnaast is er binnen dit onderzoek voor gekozen om intuïtievare benamingen te geven aan de ernstniveau's (*neutral*, *minor*, *medium* en *major*).

In de laatste fase wordt de uiteindelijke kwaliteitsscore berekend door aan elke ernstgraad een multiplicerende factor toe te kennen. Conform de richtlijnen van de MQM Foundation is het sterk aangewezen om hiervoor een exponentiële schaalvergroting te hanteren, zoals een 1-3-9 of 1-5-25 verhouding (voor respectievelijk *minor*, *medium* en *major*). Deze exponentiële weging garandeert dat zwaardere afwijkingen de kwaliteitsscore proportioneel sterker beïnvloeden. In dit onderzoek wordt de verhouding 1-3-9 toegepast zodat de score relatief gevoelig is voor zwakke variaties. Gezien de beperkte lengte van de tekstfragmenten, met statistisch gezien dus een kleiner aantal potentiële variaties, is het immers belangrijk dat een sterke variatie niet disproportioneel doorweegt. Indien de *neutral*-categorie wordt gebruikt, krijgt deze de factor 0 en heeft ze dus geen invloed op de finale score (Freitag et al., 2021; Mariana et al., 2015; MQM Foundation, 2026).

Omdat de volledige MQM-hiërarchie ('MQM Full') behoorlijk uitgebreid is, heeft de MQM Foundation ook een eenvoudiger versie uitgebracht ('MQM Core') met de belangrijkste categorieën uit de twee hoogste hiërarchische niveaus (MQM Foundation, 2026). Binnen dit onderzoek vormt MQM Full echter de basis, omdat sommige categorieën en subcategorieën van deze versie relevant zijn voor de analyse. Het hoogste niveau in de hiërarchie bevat acht categorieën: *Terminology*, *Accuracy*, *Linguistic conventions*, *Style*, *Locale conventions*, *Audience appropriateness*, *Design and markup* en *Custom*. In overeenstemming met de aanwijzingen door MQM Foundation (2026) en het gebruik van MQM in gerelateerde onderzoeken (Klubička et al., 2018; Y. Li et al., 2024; Lommel et al., 2024), wordt een selectie van relevante categorieën gebruikt voor de annotatie. Omdat dit een studie is naar 'verschuivingen' (*shifts*) en niet uitsluitend naar 'fouten' (*errors*), zijn sommige bestaande categorieën en hun beschrijving aangepast om hiermee in overeenstemming te komen. Wat volgt, is een overzicht en een beknopte verantwoording van de aangepaste MQM-typologie voor dit onderzoek, die concreter zal worden gemaakt tijdens de analyse.

- **Accuracy**
 - Mistranslation
 - Ambiguous target content
 - False friend
 - Overly literal
 - MT hallucination
 - Overtranslation
 - Undertranslation
 - Addition
 - Omission
 - Unjustified euphemism
 - Obscenity
 - Non-inclusivity
 - Untranslated
 - Culture-specific reference
 - Religious content
 - Anachronism
- **Linguistic conventions**
 - Grammar
 - Spelling
 - Duplication
 - Unintelligible
- **Style**
 - Awkward style
 - Unidiomatic style
- **Locale conventions**
- **Bracketed text**
 - Unnecessary for translation
 - Necessary for translation

De categorie *Accuracy* heeft betrekking tot de accurate representatie van inhoud uit de brontekst in de vertaling en omvat instanties in het getrouw weergeven van de originele content. De subcategorieën *Obscenity*, *Non-inclusivity*, *Culture-specific reference* en *Religious content* zijn overgeplaatst naar *Accuracy* vanuit *Audience appropriateness*, een categorie om de aanwezigheid van ongepaste inhoud voor het doelpubliek aan te duiden. Omdat deze studie net is gericht op hoe cultureel gedachtegoed wordt weergegeven, is het annoteren van deze categorie niet zinvol. De oorspronkelijke functie van sommige subcategorieën is

aangepast om conform te zijn met het onderzoek. Zo wijst *Obscenity* normaal gezien op de aanwezigheid van obscene taalgebruik in de doeltekst, maar binnen dit onderzoekskader wijst het op censuur of verbloeming van obscene inhoud uit de brontekst. De subcategorie *Anachronism* is een toevoeging om tekst te annoteren waarin iets vermeld staat dat niet bestond ten tijde van de oorspronkelijke auteur.

Verder is de functie van de categorie *Linguistic conventions* opengetrokken om niet alleen taalkundige fouten in de vertaling (onafhankelijk van de brontekst) aan te duiden, maar ook een verschoven weergave van de oorspronkelijke taalkundige structuur in de brontekst. De subcategorie *Grammar* kan dus bijvoorbeeld gebruikt worden om een grammaticale afwijking ten opzichte van de brontekst te annoteren. De categorie *Style* omvat alle formuleringen die taalkundig correct zijn maar stroef of onnatuurlijk overkomen. Indien de brontekst een stroeve vertaling onvermijdelijk maakt zonder andere opofferingen te maken, wordt er niets bij *Style* geannoteerd in dit onderzoek (in twijfelgevallen zijn de referentievertlagen doorslaggevend). Vervolgens dient *Locale conventions* om verkeerde omzettingen van variabelen aan te duiden, waaronder jaartallen, lengtematen en geldeenheden. Ten slotte is de categorie *Bracketed tekst* toegevoegd om woorden tussen ronde haken in de vertaling te annoteren. Annotaties in deze categorie krijgen soms de ernstgraad *neutral* indien het niet gaat over een verschuiving. In bijlage A wordt elke gebruikte categorie voorzien van een bondige uitleg. Met dit aangepaste MQM-framework zijn de epigrammen van Martialis, die in het volgende hoofdstuk worden besproken, geannoteerd en geanalyseerd.

3.2) Martialis' *Epigrammata*

Martialis (ca. 40 n.C. - ca. 103 n.C.) werd geboren als Romeins burger in het Spaanse provinciestedje Bilbilis, waar hij literaire scholing kreeg en woonde tot hij vroeg in de twintig was. Om carrière te maken, trok de jongeman naar Rome en presenteerde zich daar als *cliens*. De traditionele verhouding tussen *clientes* en *patroni* leverde de eerstgenoemden levensonderhoud op, in ruil voor het bijdragen aan het aanzien en imago van de *patroni*. Er is bijzonder weinig geweten over Martialis' leven in Rome vóór het uitbrengen van zijn eerste werk *Liber spectaculorum* in het jaar 80, waarna de reputatie van de dichter stelselmatig begon te groeien (Huxley, 2019; Van Dooren, 1996). In 86 verschijnen de eerste twee boeken van Martialis' meest populaire werk, een gevarieerde verzameling korte, geestige en satirische gedichten bekend onder de titel *Epigrammata*. Oorspronkelijk waren epigrammen Griekse inscripties (ἐπίγραμμα betekent letterlijk 'opschrift') op graven en votieven, maar ten tijde van Martialis was het genre reeds geëvolueerd naar korte, luchtige en persoonlijke verzen, ongeacht de drager van de tekst (Fitzgerald, 2007; Schrijvers, 2019).

De relevantie van Martialis' epigrammen binnen dit onderzoek ligt in zijn onverbloemde weergave van de Romeinse samenleving en haar culturele gedachtegoed, zoals sociale rollen, persoonlijke verhoudingen, liefde, spot, bedrog en eer. Martialis beschreef het alledaagse leven in Rome zoals het was, waardoor zijn epigrammen ook een waardevolle historische bron vormen. In diverse contexten spelen zakenlui, erfenisjagers, ouderen, pederasten en nog vele andere typen in zijn werk een rol in het creëren van een herkenbaar realisme. Het maakt niet uit welke sociale status, gender of positie iemand had, iedereen wordt bekeken met dezelfde geamuseerde, ironische blik (Schrijvers, 2019; Van Dooren, 1996). Een uitzondering hierop is de vleierij voor keizer Domitianus op een niveau dat voor hedendaagse lezers hoogst onnatuurlijk aanvoelt. Deze lofprijzingen waren waarschijnlijk geworteld in het Romeinse principe van *do ut des* ('Ik geef opdat jij geeft.'), een gedragsnorm die Martialis talrijke keizerlijke gunsten en invitaties heeft opgeleverd in ruil voor schriftelijke lofprijzingen (Huxley, 2019; Van Dooren, 1996).

Martialis had aandacht voor de ongewone dingen in de grootstad, waaronder allerlei perversiteiten, en dreef dikwijls de spot met alle vormen van intimiteit en de beoefenaars ervan. Het merendeel van zijn werk behandelt echter andere thema's zoals satire op persoonlijke eigenschappen, beschrijvingen van villa's en huwelijkswensen. De epigrammen zijn niet thematisch of narratief geordend, met als doel om verveling bij de lezer te beperken en de grote variatie in metrum, lengte en onderwerp duidelijk zichtbaar te maken (Schrijvers, 2019; Van Dooren, 1996). De stijl van Martialis' epigrammen wordt gekenmerkt door gevatheid, ironie, vormelijke perfectie en de inclusie van een pointe of ontknoping op het einde (Held, 2016; Van Dooren, 1996). De bondigheid en ironische toon vormen een uitdagende setting waarin vertalers regelmatig moeten kiezen tussen vlotheid of een getrouwe weergave bieden van de brontekst. Deze vertaalkeuze is vergelijkbaar met het onderscheid tussen *domestication* en *foreignisation* (Venuti, 2017), zoals eerder besproken in het hoofdstuk 'Een vertaalwetenschappelijk perspectief'.

Eenzijds was de Romeinse auteur reeds in zijn eigen leven erg populair bij het brede publiek en zou hij later een voorbeeld zijn voor onder meer Erasmus, Voltaire en Goethe (Van Dooren, 1996). Anderzijds zorgde het obscene, pornografische karakter van sommige epigrammen voor afkeuring en verwerping bij puriteinen. Hiervan getuigen meerdere 'opgeschoonde' Latijnse edities, die *obscaenis demptis* ('met weglating van de obscene gedichten') op de voorpagina hebben staan. Deze trend zette zich voort in de latere receptie, in de vorm van zorgvuldig geselecteerde en moreel verantwoorde vertalingen. Het verchristelijken van Martialis, zoals Maarten Luther in de zestiende eeuw deed in een vertaling van epigram 10.47, was een manier om de inhoud van bepaalde epigrammen acceptabel te maken, alsook het maskeren

van ‘controversiële’ (meestal homoseksuele) seksuele gedragingen (Schrijvers, 2019; Van Dooren, 1996). Schrijvers (2019) beargumenteert dat de gemengde, tegenstrijdige receptie van Martialis’ poëzie aan de basis ligt van het geringe aantal Nederlandse vertalingen, zoals besproken wordt in het volgende hoofdstuk.

3.3) Keuze van professionele vertalingen

De vertalingen door de LLM's worden vergeleken met menselijke, gepubliceerde vertalingen, die dienen als referentie gedurende de annotatie en de bespreking van resultaten. Het doel hiervan is om een kader te scheppen waartegen de LLM-output kan worden afgewogen. Door te steunen op meerdere vakkundige referentieververtalingen, wordt bovendien de invloed van subjectieve interpretaties en individuele voorkeuren tijdens het onderzoek geminimaliseerd. De geselecteerde vertalingen zijn ‘Martialis, *Verzamelde epigrammen*: Vertaald, ingeleid en toegelicht door Piet Schrijvers’ (2019) en ‘Martialis, *Romeinse epigrammen*: Vertaald en ingeleid door Frans Van Dooren’ (1996). Beide vertalingen baseren zich op de Latijn-Engelse Loeb-editie, respectievelijk uit 1993 (Martial, 1993) en 1947 (Martial, 1947). Er is gekozen voor Nederlandse vertalingen vanwege het kleine aantal moderne Martialisvertalingen die beschikbaar zijn in de Nederlandse taal (Schrijvers, 2019; Van Houdt, 2020), waardoor gegenereerde vertalingen een snel en gemakkelijk alternatief vormen.

De metrische vertaling door Piet Schrijvers dateert uit 2019 en is daarmee momenteel de meest recente Nederlandse vertaling van Martialis’ epigrammen met enige bekendheid. Het is ook de eerste Nederlandse vertaling van de volledige verzameling Martialisepigrammen, boeken één tot en met twaalf, voorzien van een korte introductie en voetnoten. In zijn vertaling en parateksten toont Schrijvers, ‘die zijn lauweren intussen meer dan verdiend heeft met het vertalen van klassieke Latijnse poëzie’ (Van Houdt, 2020, p. 1) een contextbewuste en deskundige visie op Martialis en Martialisvertalingen. Schrijvers (2020) stelt bovendien in het tijdschrift *Filter* dat zijn voorkeur uitgaat naar een ‘foreignizing’ vertaalstijl met ‘zo veel mogelijk handhaving van de oorspronkelijke vorm en inhoud’ (p. 61), wat interessant is binnen dit onderzoekskader. Classicus en vertaler Patrick De Rynck schrijft op gelijkaardige wijze: ‘Schrijvers wil als Martialisvertaler de eerste foreignizer zijn en hij behoudt daarom de oorspronkelijke vorm van de Latijnse gedichten.’ (Patrick De Rynck, 2019). Verder wordt de vertaling van Schrijvers gekenmerkt door helder en levendig taalgebruik, waaruit de kwaliteiten van Martialis’ epigrammen zich duidelijk tonen.

De vertaling door Frans van Dooren (1996) is een bloemlezing van 250 gedichten uit de *Epigrammata*, aangevuld met een onderlegde inleiding en verantwoording. Zijn selectie van epigrammen is uitgevoerd met enige subjectiviteit, maar met de voorzorg

dat ‘alle aspecten van Martialis’ poëzie zo evenwichtig mogelijk aan bod komen’ (Van Dooren, 1996, p. 155). Piet Schrijvers (2019) becommentarieert tevens dat Van Dooren de gevarieerde dosering heeft gehandhaafd die Martialis nastreefde. Deze studie tracht die variatie na te volgen door de epigrammenselectie te beperken tot Van Doorens bloemlezing. Op die manier is er ook minder risico op *cherrypicking*. Omdat deze vertaling minder recent is en bovendien consequent eindrijm inlast, wordt ze eerder als indicatieve referentie gebruikt dan dat er een letterlijke overeenkomst moet zijn met de gegenereerde vertalingen. Alle referentieververtalingen zijn te vinden in bijlagen H en I.

3.4) Uitvoering

De analyse van Martialisvertalingen door LLM's in dit onderzoek bestaat uit drie stappen: het genereren van de vertalingen, het annoteren en ten slotte een gedetailleerde bespreking van de meest relevante verschuivingen. De LLM-vertalingen zijn gegenereerd met een eenvoudige *zero-shot*-prompt: ‘Vertaal de volgende tekst: [...]’, waarbij [...] vervangen werd door de Latijnse tekst. Het doel is om de repliceerbaarheid in verder onderzoek te vergroten en om de invloed van de specifieke prompt minimaal te houden. In verschillende studies wordt immers betoogd dat LLM's gevoelig zijn voor semantisch-identieke variaties in een prompt. Deze ‘Prompt Sensitivity’ zorgt ervoor dat stilistische verschillen in de prompt – zelfs het toevoegen van een extra spatie – een helemaal andere output kunnen veroorzaken (He et al., 2024; Ngweta et al. 2025). Zoals besproken in de literatuurstudie, is de aangewezen *prompting*-strategie bovendien verschoven naar korte, duidelijke instructies door de opkomst van *reasoning*-LLM's (Lee et al., 2025; X. Li et al., 2025). De inherente redeneerstappen van recentere modellen verlagen de invloed die uitgebreide prompts hebben op de output en kunnen zelfs inefficiëntie of fouten veroorzaken indien de prompt niet optimaal is opgesteld (Wu, 2025). Gemini 3.1 pro, M365 Copilot Chat en Kimi-K2.6 gebruiken allemaal intrinsiek redeneren (Davis, 2026; The Gemini Team, 2026; Kimi Team et al., 2026). Om de neutraliteit van de resultaten te vergroten, is de optie om gepersonaliseerde data uit voorgaande chats te gebruiken bij elk model uitgeschakeld en wordt elke prompt in een nieuw chatvenster gelanceerd.

De Latijnse tekst in de prompts bestaat telkens uit één integraal epigram van Martialis, overgenomen uit de Latijn-Engelse Loeb-editie (Martial, 1993). De behandelde epigrammen zijn 1.1, 1.4, 1.16, 1.19, 1.23, 1.24, 1.28, 1.32, 1.33, 1.34, 2.3, 2.7, 2.8, 2.11 en 2.13. Dit zijn de eerste tien epigrammen uit de bloemlezing door Frans Van Dooren, zoals besproken in het hoofdstuk ‘Keuze van professionele vertalingen’, aangevuld met de eerste vijf epigrammen uit de bloemlezing van het tweede boek. De Latijnse epigrammen zijn terug te vinden in bijlage G. Elke vertaling wordt slechts eenmalig gegenereerd en vervolgens gekopieerd naar de ‘MQM annotation app’.

Deze applicatie is via Gemini Canvas (Google, z.d.) ontwikkeld als gebruiksvriendelijke annotatietool voor dit onderzoek, maar is ook bruikbaar voor andere MQM-evaluaties. De annotatiecategorïeën zijn immers volledig aanpasbaar en kunnen nadien als een apart bestand worden opgeslagen. De applicatie laadt standaard in met de volledige typologie van MQM Full. Het annoteren en evalueren van een tekstfragment verloopt volgens de richtlijnen van het MQM-framework, maar is minder tijdrovend ten opzichte van alternatieven zoals Excel. Daarnaast wordt voor elke annotatiecategorie een uitleg voorzien, die (net zoals de categorie zelf) aanpasbaar is indien nodig. De praktische werking van de applicatie staat verder uitgelegd in bijlage B en in de app zelf, waarvan de HTML-code gegeven is in bijlage C.

Informatie in de LLM-output die niet behoort tot de vertaling zelf, wordt niet mee gekopieerd of geannoteerd. Uiteraard is niet elke vorm van afwijking sterk genoeg om een ‘verschuiving’ genoemd te worden. Sommige vertalingen zijn verschillend van elkaar op een letterlijk niveau, maar zijn semantisch, pragmatisch en stilistisch nagenoeg volledig equivalent aan elkaar. Een klein verschil in woordvolgorde, of het vertalen van een bescheidenheidsmeervoud (*Pluralis modestiae*) als ‘wij’ of ‘ik’, is zodanig triviaal bij Latijnse vertalingen dat het niet wordt geannoteerd. Daarnaast genereert Copilot in de output voor epigrammen 1.1, 1.16, 1.23, 1.24, 1.28, 1.32, 1.33, 2.3 en 2.13 zowel een ‘vloeiendere’ of ‘vlottere’ vertaling als een meer ‘letterlijke’ vertaling. Na het annoteren en vergelijken van beide vertalingen voor epigrammen 1.1, 1.16 en 1.23, is ervoor geopteerd om enkel de ‘vlottere’ vertalingen te betrekken bij de analyse. De ‘letterlijkere’ vertalingen vertonen immers ongeveer dezelfde verschuivingen als de ‘vlottere’ vertalingen, maar dan met minder heldere zinsbouw en meer tekst tussen haakjes. De resultaten voor de drie geannoteerde, ‘letterlijkere’ vertalingen staan in bijlage M. Kimi genereert ook twee vertalingen voor epigram 1.16, maar beide vertonen geen verschuivingen ten opzichte van de brontekst. Een overzicht van alle geannoteerde verschuivingen en de LLM-vertalingen zelf is terug te vinden in de bijlagen J, K en L, waarvan de belangrijkste resultaten en tendensen in het volgende hoofdstuk worden gepresenteerd.

4) Analyse

Door gebruik te maken van MQM-annotaties kunnen de verschuivingen tussen de bron- en doelttekst voor elk model nauwkeurig in kaart worden gebracht en gekwantificeerd. Tabel 1 biedt een overzicht van de annotaties per categorie. De punten van ondergeschikte categorieën tellen automatisch bij het totaal van de overkoepelende categorie. Het is echter ook mogelijk om een annotatie uit te voeren in die hogere categorie zelf. In dit onderzoek staat steeds aangegeven wanneer de punten van een rechtstreekse annotatie zijn weergegeven en wanneer de totaalsom. Gedetailleerdere statistieken van de verschillen tussen de modellen zijn terug te vinden in bijlage D. In de volgende alinea's worden de meest opvallende bevindingen besproken en geïllustreerd met voorbeelden, om de manier waarop LLM-vertalingen cultureel gedachtegoed weergeven te concretiseren.

Categorie	Puntentotaal	Aantal shifts	Minor shifts	Medium shifts	Major shifts	Neutral shifts
Totaal van alle categorieën	262	118	67	37	9	5
ACCURACY	193 (specifiek: 8)	86 (specifiek: 10)	8	0	0	2
Mistranslation	94 (specifiek: 42)	34 (specifiek: 12)	6	3	3	0
<i>Ambiguous target content</i>	5	3	2	1	0	0
<i>False friend</i>	3	1	0	1	0	0
<i>Overly literal</i>	44	18	8	9	1	0
<i>MT hallucination</i>	0	0	0	0	0	0
Overtranslation	5	5	5	0	0	0
Undertranslation	19	9	4	5	0	0
Addition	18	14	12	2	0	0
Omission	14	4	2	1	1	0
Unjustified euphemism	10 (specifiek: 0)	2 (specifiek: 0)	0	0	0	0
<i>Obscenity</i>	9	1	0	0	1	0
<i>Non-inclusivity</i>	1	1	1	0	0	0
Untranslated	15	4	0	2	1	1
Culture-specific reference	10	4	1	3	0	0
<i>Religious content</i>	0	0	0	0	0	0
Anachronism	0	0	0	0	0	0
LINGUISTIC CONVENTIONS	33 (specifiek: 0)	17 (specifiek: 0)	0	0	0	0
Grammar	30	16	9	7	0	0
Spelling	0	0	0	0	0	0
Duplication	3	1	0	1	0	0
Unintelligible	0	0	0	0	0	0

STYLE	30 (specifiek: 23)	12 (specifiek: 7)	5	0	2	0
Awkward style	0	0	0	0	0	0
Unidiomatic style	7	5	4	1	0	0
LOCALE CONVENTIONS	0	0	0	0	0	0
BRACKETED TEXT	3 (specifiek: 0)	3 (specifiek: 0)	0	0	0	0
Unnecessary for translation	3	2	0	1	0	1
Necessary for translation	0	1	0	0	0	1

Tabel 1: Overzicht van de annotaties per categorie.

4.1) Belangrijkste verschuivingen in de LLM-vertalingen

4.1.1) Overly literal shifts

De subcategorie met het hoogste aantal punten is *overly literal* (in totaal 44 punten), wat een indicatie is van het frequente gebruik van letterlijkere vertalingen door de LLM's. Verschuivingen behoren tot deze categorie wanneer een vertaling de figuurlijke of afgeleide betekenis in de brontekst niet overneemt, maar in de plaats daarvan een letterlijke of meer gebruikelijke weergave kiest. Omdat de specifiekere betekenis verloren gaat in de vertaling, verandert of verduistert de boodschap van de tekst en haar culturele ideeëngoed.

Overly literal shifts	Totaal	Minor	Medium	Major
Kimi	9	4	4	1
Copilot	8	3	5	0
Gemini	1	1	0	0
Alle modellen	18	8	9	1

Tabel 2: Overzicht van overly literal shifts.

Brontekst [Ep. 1.34] (Martial, 1993)	<i>Incustoditis et apertis, Lesbia, semper liminibus peccas nec tua furta tegis</i>
Referentievertaling (Schrijvers, 2019)	Altijd is jouw deur onbewaakt en staat open, Lesbia, als je vreemdgaat en niet je overspel bedekt.
Vertaling door LLM (Kimi)	Met onbewaakte en wijdopen deuren, Lesbia, bega je voortdurend overspel en verberg je je diefstallen niet.

Tabel 3: Voorbeeld bij overly literal shifts (1).

Zo is de vertaling ‘diefstallen’ voor *furta* in epigram 1.34 (zie Tabel 3) door Kimi een *medium overly literal* shift, aangezien *furtum* overdrachtelijk ook een andere heimelijke handeling, zoals ‘overspel’ kan betekenen (Pinkster, 2012d). Hier gebruikt Martialis deze term om het gedrag van Lesbia te beschrijven in een seksuele context, dus zorgt de basisvertaling ‘diefstallen’ voor een verschoven betekenis.

Brontekst [Ep. 1.34] (Martial, 1993)	<i>at meretrix abigit testem veloque seraque raraque Summemmi fornice rima patet.</i>
Referentieverting (Schrijvers, 2019)	Maar een prostituee verwijdt met gordijn en grendel iedere getuige, geen spleet staat open in een bordeel
Vertaling door LLM (Kimi)	Maar zelfs een hoer jaagt de getuige weg, en met gordijn en slot — slechts zelden staat er een spleet open in de boog van Summemmus.

Tabel 4: Voorbeeld bij overly literal shifts (2).

Verderop wordt *fornice* vertaald naar ‘de boog’ (zie Tabel 4). Een *fornix* slaat oorspronkelijk op een ‘gewelf’ of ‘(triomf)boog’, maar de betekenis ervan kan ook worden opengetrokken naar ‘onderaards gewelf’, ‘kelder’ en ook ‘bordeel’. Gezien de context van prostitutie en de vermelding van *Summemmus*, een veronderstelde bordelenwijk in Rome (Martial, 1993), gaat het om deze laatste, figuurlijke betekenis en wordt een *major* shift geannoteerd. De historische verwijzing naar *Summemmus* gaat overigens ook verloren door deze verschuiving.

Brontekst [Ep. 1.23] (Martial, 1993)	<i>Invitas nullum nisi cum quo, Cotta, lavis et dant convivam balnea sola tibi.</i>
Referentieverting (Schrijvers, 2019)	Cotta, jij nodigt slechts mensen uit met wie je kunt baden. Louer het badhuis levert aan jou een gast.
Vertaling door LLM (Kimi)	Je nodigt nooit iemand uit, Cotta, tenzij met wie je badt, en alleen de baden geven jou een gast.

Tabel 5: Voorbeeld bij overly literal shifts (3).

Zoals blijkt uit Tabel 5, treedt er vervolgens bij Kimi nog een relevante *minor overly literal* shift op in de vertaling ‘baden’, voor *balnea* in epigram 1.23. Een *balneum*, het enkelvoud van *balnea*, draagt de betekenis van ‘bad(kuip)’ of ‘badkamer’, maar het meervoud wordt gebruikt om te verwijzen naar de thermen (Pinkster, 2012a). Schrijvers en van Dooren vertalen respectievelijk ook met ‘badhuis’ en ‘thermen’. Deze

verschuiving is een illustratie van hoe zelfs een kleine afwijking ten opzichte van de brontekst schadelijk kan zijn voor de weergave van cultureel gedachtegoed.

Brontekst [Ep. 2.8] (Martial, 1993)	<i>quod si non illum sed me peccasse putabis, tunc ego te credam cordis habere nihil.</i>
Referentievertlaling (Schrijvers, 2019)	Als je denkt dat niet hij, maar dat ik in de fout was, zal ik denken dat je er geen verstand van hebt.
Vertaling door LLM (Copilot)	Maar als je denkt dat niet hij, maar ik heb gezondigd , dan zal ik geloven dat jij geen hart hebt.

Tabel 6: Voorbeeld bij overly literal shifts (4).

In epigram 2.8 wordt het werkwoord *peccasse* in de zin *quod si non illum sed me peccasse putabis* door Copilot vertaald als ‘gezondigd’ (zie Tabel 6), wat een *medium* shift is. *Peccare* kan immers ook ‘een fout maken’ betekenen, zonder de morele lading van de basisvertaling ‘zondigen’ (Pinkster, 2012g). Het gaat hier over taalfouten maken in een epigrammenbundel, dus er is geen uitgesproken morele connotatie verbonden aan *peccasse*. In het volgende vers van dat epigram vertaalt Copilot het Latijnse *cordis* als ‘hart’ in de zin ‘dan zal ik geloven dat jij geen hart hebt’ (Tabel 6), terwijl *cor* ook op ‘verstand’ of ‘inzicht’ kan wijzen in deze narratieve context (Pinkster, 2012b). Op basis van de vertalingen door Schrijvers en Van Dooren wordt een *medium* shift geannoteerd. Kimi gebruikt overigens ook de woorden ‘gezondigd’ en ‘hart’ in de vertaling, met dezelfde verschuivingen als gevolg.

Brontekst [Ep. 1.24] (Martial, 1993)	<i>Aspicias incompitis illum, Deciane, capillis, cuius et ipse times triste supercilium</i>
Referentievertlaling (Schrijvers, 2019)	Zie je die man, Decianus, met die verwaarloosde haren, door wiens grimmige frons jij wordt geïmponeerd
Vertaling door LLM (Copilot)	Je ziet hem daar, Decianus, met zijn onverzorgde haar— die man van wie zelfs jij die sombere frons vreest

Tabel 7: Voorbeeld bij overly literal shifts (5).

Ten slotte is de vertaling ‘sombere frons’ voor *triste supercilium* in epigram 1.24 door Copilot een *medium* shift (zie Tabel 7). Het werkwoord *times* (‘jij vreest’) geeft immers aan dat *triste* hier eerder in de betekenis van ‘streng’ of angstaanjagend moet worden geïnterpreteerd. Deze verschuiving schept een vertekend beeld door het culturele

stereotiep van een mannelijke, serieuze figuur die zich niet om zijn haartooi bekommert weer te geven als een sombere slons.

4.1.2) Undertranslation shifts

De geannoteerde *undertranslation* shifts zijn afgezwakte of minder specifieke weergaven van de oorspronkelijke inhoud. Veel verschuivingen in deze categorie komen neer op algemener woordgebruik dat een afvlakking teweegbrengt van cultureel ideeëngoed in de vertaling.

<u>Undertranslation shifts</u>	Totaal	Minor	Medium	Major
Kimi	3	1	2	0
Copilot	5	2	3	0
Gemini	1	1	0	0
Alle modellen	9	4	5	0

Tabel 8: Overzicht van undertranslation shifts.

Brontekst [Ep. 1.34] (Martial, 1993)

*A Chione saltem vel ab lade disce pudorem:
 abscondunt spurcas et **monumenta** lupas.*

Referentievertlaling (Schrijvers, 2019)

Leer toch tenminste van Chia of las gevoelens van schaamte: grafmonumenten verbergen de allersmerigste hoer.

Vertaling door LLM (Kimi)

Leer tenminste van Chione of van las wat schaamte is: zelfs **monumenten** verbergen die vuile wolvin-nachten.

Tabel 9: Voorbeeld bij undertranslation shifts (1).

De eerste relevante verschuiving in deze categorie bevindt zich in de vertaling door Kimi van epigram 1.34 (zie Tabel 9), waarin Martialis verwijst naar de armste laag van prostituees, die vaak hun diensten aanboden op begraafplaatsen en tussen grafmonumenten om verborgen te blijven. De vertaling ‘monumenten’ voor *monumenta* is dus een *medium undertranslation* shift, aangezien ze een minder specifieke weergave biedt van dit culturele fenomeen dan de brontekst.

Brontekst [Ep. 1.24] (Martial, 1993)	<i>Aspicias incomptis illum, Deciane, capillis, cuius et ipse times triste supercilium, qui loquitur Curios adsertoresque Camillos? nolito fronti credere: nupsit heri</i>
Referentieverting (Schrijvers, 2019)	Zie je die man, Decianus, met die verwaarloosde haren, door wiens grimmige frons jij wordt geïmponeerd, die zijn mond vol heeft van Curius en andere vrijheidsstrijders? Vergis je niet in die aanblik. Gisteren was hij de bruid!
Vertaling door LLM (Copilot)	Je ziet hem daar, Decianus, met zijn onverzorgde haar— die man van wie zelfs jij die sombere frons vreest, die maar blijft praten over de Curii en de Camilli, die deugdhelden. Geloof zijn voorhoofd niet: gisteren is hij getrouwd .

Tabel 10: Voorbeeld bij undertranslation shifts (2).

Het werkwoord *nupsit* in het laatste vers van epigram 1.24, de pointe van dit epigram, vertaalde Copilot als ‘is hij getrouwd’ (zie Tabel 10). Hierdoor gaat een cruciaal onderdeel van de grap verloren, omdat *nubere* haast uitsluitend gebruikt werd om te verwijzen naar het ‘sluieren’ en dus het ‘trouwen’ van specifiek de bruid. De schijnbaar stoere en traditionele man nam dus onlangs een vrouwelijke, passieve rol aan bij het huwelijk. Wanneer een man trouwde, gebruikte men doorgaans *ducere*, wat de actieve en ‘leidende’ rol van de bruidegom in Romeinse huwelijkspraktijken symboliseert. Het onderscheid tussen een actieve of passieve rol in een relatie of huwelijk was diep verweven met genderrollen in de Oudheid en werd belangrijker geacht dan het biologische geslacht van de partners (Eyben et al., 2003) Deze culturele achtergrond construeert de betekenis en humor van het gedicht, maar is niet zichtbaar in de vertaling door Copilot.

4.1.3) Overtranslation shifts

Overtranslation shifts zijn verschuivingen waarbij de vertaling inhoud uit de brontekst explicieter en duidelijker maakt voor het doelpubliek. Voor cultureel gedachtegoed betekent dit doorgaans een specifiekere invulling van een bepaald concept uit de oorspronkelijke tekst.

Overtranslation shifts	Totaal	Minor	Medium	Major
Kimi	2	2	0	0
Copilot	0	0	0	0
Gemini	3	3	0	0
Alle modellen	5	5	0	0

Tabel 11: Overzicht van overtranslation shifts.

Brontekst [Ep. 1.34] (Martial, 1993)	<i>et plus spectator quam te delectat adulter nec sunt grata tibi gaudia si qua latent.</i>
Referentieverting (Schrijvers, 2019)	Je hebt meer plezier in een toeschouwer dan in een minnaar, genoegens bevallen jou niet als ze verborgen zijn.
Vertaling door LLM (Kimi)	Een toeschouwer bevalt je meer dan de overspelige man zelf, en genoegens die ergens verborgen blijven, zijn je niet welgevallig.

Tabel 12: Voorbeeld bij overtranslation shifts (1).

Een voorbeeld hiervan treedt op in de vertaling ‘overspelige man’ door Kimi bij epigram 1.34, zoals blijkt uit Tabel 12. Het woord *adulter* specificeert immers niet het geslacht van de overspelige (Andrews, 1879). Deze *overtranslation* shift komt overigens ook voor bij de referentieverting door Schrijvers, aangezien ‘minnaar’ een mannelijk geslacht suggereert.

Er zijn in totaal slechts vijf annotaties uitgevoerd in deze categorie, wat suggereert dat dit soort verschuiving relatief weinig voorkomt in LLM-vertalingen. Tabel 13 toont bovendien aan dat, hoewel Gemini iets meer *overtranslation* shifts veroorzaakt dan *undertranslation* shifts, bij Kimi en Copilot de hoeveelheid en ernstgraad van de verschuivingen beduidend hoger zijn voor de categorie *undertranslation*.

	Kimi	Copilot	Gemini	Totaal
Overtransl. (pt.)	2	0	3	5
Overtransl. (shifts)	2 min	0	3 min	5 min
Undertransl. (pt.)	7	11	1	19
Undertransl. (shifts)	1 min, 2 med	2 min, 3 med	1 min	4 min, 5 med

Tabel 13: Vergelijking van de annotatiecategorieën overtranslation en undertranslation.

4.1.4) Addition

De annotaties met categorie *addition* zijn directe toevoegingen in de vertaling die niet in de brontekst voorkomen. De verschuivingen betreffen voornamelijk korte woorden die nadruk leggen om de boodschap te expliciteren. Deze toevoegingen hebben een assimilerend effect door culturele inhoud toegankelijker te maken voor het doelpubliek.

<u>Addition shifts</u>	Totaal	Minor	Medium	Major
Kimi	2	2	0	0
Copilot	5	3	2	0
Gemini	7	7	0	0
Alle modellen	14	12	2	0

Tabel 14: Overzicht van addition shifts.

Brontekst [Ep. 1.34] (Martial, 1993)

at meretrix abigit testem veloce seraque

Referentieverting (Schrijvers, 2019)

Maar een prostituee verwijdt met gordijn en grendel iedere getuige

Vertaling door LLM (Kimi)

Maar **zelfs** een hoer jaagt de getuige weg, en met gordijn en slot

Tabel 15: Voorbeeld bij addition shifts (1).

Zo voegt Kimi in de vertaling van epigram 1.34 het woord ‘zelfs’ toe in de zin ‘maar zelfs een hoer jaagt de getuige weg’ (zie Tabel 15). Door deze aanvulling wordt het contrast explicieter gemaakt tussen Lesbia, wier gedrag Martialis in dit epigram bespot, en ‘normale’ prostituees.

Brontekst [Ep. 1.24] (Martial, 1993)

*qui loquitur **Curios** adsertoresque **Camillos**?*

Referentieverting (Schrijvers, 2019)

die zijn mond vol heeft van Curius en andere vrijheidsstrijders?

Vertaling door LLM (Gemini)

die de mond vol heeft van **(oer-Romeinse helden als)** de Curii en de bevrijders, de Camilli?

Tabel 16: Voorbeeld bij addition shifts (2).

Daarnaast geeft Gemini in epigram 1.24 uitleg bij de twee historische verwijzingen *Curios* en *Camillos* door ‘(oer-Romeinse helden als)’ toe te voegen in de vertaling, zoals Tabel 16 weergeeft. Deze informatie wordt wel geïmpliceerd met de vermelding van de twee personen Curius en Camillus, die symbool stonden voor traditionele Romeinse waarden (Snicket, 2006), maar staat niet in de brontekst. Door de achterliggende gedachte van de eigennamen te expliciteren, verminderen de mogelijke associaties bij deze cultuurgebonden verwijzing, wat de beeldende kracht beïnvloedt. De verschuiving is geannoteerd als een *medium unnecessary for translation* shift in de hoofdcategorie *bracketed tekst*, aangezien ze tussen haakjes staat, maar ze veroorzaakt in feite hetzelfde effect als een *addition* shift. *Bracketed text* en *addition* zijn overigens de enige categorieën waarin Gemini een groter aantal

shifts heeft dan Copilot en Kimi. Hieruit blijkt dat Gemini in verhouding meer assimilerende verschuivingen veroorzaakte dan de andere twee modellen.

4.1.5) Omission

Voor *omission*, een categorie die het weglaten van inhoud uit de brontekst aanduidt, zijn er twee *minor* shifts, een *medium* shift en een *major* shift, waarbij deze laatste werd aangeduid omdat Copilot weigerde epigram 1.34 te vertalen wegens expliciet seksuele inhoud (cf. infra).

<u>Addition shifts</u>	Totaal	Minor	Medium	Major
Kimi	0	0	0	0
Copilot	3	1	1	1
Gemini	1	1	0	0
Alle modellen	4	2	1	1

Tabel 17: Overzicht van omission shifts.

Een vergelijking van de annotatiecategorieën *addition* met veertien verschuivingen en *omission* met vier verschuivingen (zie Tabel 18), toont aan dat er in verhouding frequenter stukjes tekst worden toegevoegd in de LLM-vertalingen dan dat er inhoud volledig ontbreekt.

	Kimi	Copilot	Gemini	Totaal
Addition (pt.)	2	9	7	18
Addition (shifts)	2 min	3 min, 2 med	7 min	12 min, 2 med
Omission (pt.)	0	13	1	14
Omission (shifts)	0	1 min, 1 med, 1 maj	1 min	2 min, 1 med, 1 maj

Tabel 18: Vergelijking van de annotatiecategorieën addition en omission.

4.1.6) Mistranslation shifts

Een andere omvangrijke categorie in dit onderzoek is *mistranslation* met 42 rechtstreekse annotatiepunten. Algemene semantische afwijkingen die niet bij een andere categorie passen, zijn geannoteerd als *mistranslation* shifts. Het grootste aandeel binnen deze categorie bevindt zich in enkele *major* verschuivingen die de betekenis van cultuurgebonden inhoud grondig wijzigen.

Mistranslation shifts	Totaal	Minor	Medium	Major
Kimi	6	3	1	2
Copilot	4	2	1	1
Gemini	2	1	1	0
Alle modellen	12	6	3	3

Tabel 19: Overzicht van mistranslation shifts.

Brontekst [Ep. 1.4] (Martial, 1993)

*qua Thymelen spectas derisoremque **Latinum**,
 illa fronte precor carmina nostra legas*

Referentieverting (Schrijvers, 2019)

Met het gezicht waarmee u kijkt naar ballet of
 een mime,
 moet u mijn epigrammen lezen, zo bid ik u.

Vertaling door LLM (Kimi)

Waar je naar Thymele en de **Latijnse**
 spotvogel kijkt,
 daarop, bid ik, lees mijn gedichten met
 diezelfde blik.

Tabel 20: Voorbeeld bij mistranslation shifts (1).

Een duidelijke verschuiving treedt op in epigram 1.4, waar *Latinum* door Kimi wordt weergegeven als het bijvoeglijk naamwoord ‘Latijnse’ (zie Tabel 20). *Latinus* is echter ook de eigennaam van een bekende mimespeler uit de tijd van Domitianus (Gogräfe et al., z.d.). Omdat de verwijzing naar een culturele en historische figuur verloren gaat, krijgt deze shift een *major* ernstgraad toegewezen. In de vertalingen door Schrijvers en Van Dooren wordt de mimespeler overigens niet bij naam genoemd, maar blijft de verwijzing naar het beroep wel behouden. Hoewel de afwijking bij Kimi’s vertaling impactvoller is dan bij de referentievertingen, veroorzaken beide een verschoven weergave van cultureel gedachtegoed.

Brontekst [Ep. 1.34] (Martial, 1993)

*A Chione saltem vel ab lade disce pudorem:
 abscondunt spurcas et monumenta **lupas**.*

Referentieverting (Schrijvers, 2019)

Leer toch tenminste van Chia of las gevoelens
 van schaamte:
 grafmonumenten verbergen de allersmerigste
 hoer.

Vertaling door LLM (Kimi)

Leer tenminste van Chione of van las wat
 schaamte is:
 zelfs monumenten verbergen die vuile **wolvin-
 nachten**.

Tabel 21: Voorbeeld bij mistranslation shifts (2).

Daarnaast wordt in epigram 1.34 *lupas* vertaald als ‘wolvijn-nachten’ door Kimi (zie Tabel 21), wellicht gebaseerd op de betekenis ‘wolvijn’, die *lupa* in de eerste plaats draagt. Het woord duidt echter ook prostituees aan op een pejoratieve manier (Pinkster, 2012e). Net zoals bij de eerder besproken *overly literal* shifts die Kimi in dit epigram veroorzaakt, lijkt de seksuele context hier problematisch te zijn voor de LLM-vertaling. De verschuiving is geannoteerd als een *major mistranslation*.

4.1.7) Culture-specific reference shifts

Deze categorie omvat afwijkende weergaven van cultureel gedachtegoed waarbij de vertaling wel in hetzelfde semantisch veld blijft als de brontekst, maar er binnen dat veld toch een verschuiving plaatsvindt.

Culture-specific reference shifts	Totaal	Minor	Medium	Major
Kimi	2	1	1	0
Copilot	1	0	1	0
Gemini	1	0	1	0
Alle modellen	4	1	3	0

Tabel 22: Overzicht van culture-specific reference shifts.

Brontekst [Ep. 1.1] (Martial, 1993)	<i>cui, lector studiose, quod dedisti viventibus decus atque sentienti, rari post cineres habent poetae.</i>
Referentievertlaling (Schrijvers, 2019)	Beste lezer, de roem die je hem verleende bij zijn leven en bewustzijn, krijgen dichters zelden als ze gecremeerd zijn.
Vertaling door LLM (Kimi)	aan wie jij, ijverige lezer, wat je ook gaf aan eer terwijl hij leefde en voelde, dat hebben zeldzame dichters na hun as.

Tabel 23: Voorbeeld bij culture-specific reference shifts (1).

Zo veroorzaakt Kimi met de vertaling ‘eer’ voor het woord *decus* in epigram 1.1 (zie Tabel 23) een verschuiving in culturele weergave. In de Romeinse samenleving werd er immers onderscheid gemaakt tussen *honor* ofwel ‘eer’, de maatschappelijke erkenning voor geleverde prestaties – zoals een verkiezing winnen of een triomftocht ontvangen – , *fama* ofwel ‘reputatie’, wat anderen over je zeggen tijdens je leven en *gloria* ofwel ‘roem’, wat anderen over je zeggen tijdens je leven en in de loop der geschiedenis. In dit geval gaat het specifiek over ‘roem’, aangezien Martialis schrijft dat dichters het ook postuum kunnen verkrijgen. Omdat *decus* in een neutrale context zowel ‘eer’ als ‘roem’ kan betekenen (Pinkster, 2012c) en de lijn tussen deze vertalingen in het Nederlands nogal dun is, is er gekozen voor de ernstgraad *minor*.

Brontekst [Ep. 1.4] (Martial, 1993)	<i>innocuos censura potest permittere lusus: lasciva est nobis pagina, vita proba.</i>
Referentievertaling (Schrijvers, 2019)	Een censor kan aan onschuldige scherts zijn toestemming geven. Mijn pagina is gewaagd, mijn levenswijze fatsoenlijk.
Vertaling door LLM (Kimi)	De censuur kan onschuldige spelletjes toestaan: onze pagina is wellustig, ons leven is deugdzaam.
Vertaling door LLM (Copilot)	Want de censuur mag onschuldige spelerei toelaten: mijn pagina is vrijpostig, maar mijn leven is deugdzaam.

Tabel 24: Voorbeeld bij culture-specific reference shifts (2).

Verder vertalen zowel Kimi als Copilot het woord *censura* in epigram 1.4 als ‘censuur’ (zie Tabel 24), terwijl de betekenis eigenlijk slaat op een Romeinse ambtspositie. Een *censor* was immers iemand die controle uitvoerde op – onder meer – openbare zeden, maar het Nederlandse woord ‘censuur’ wordt niet gebruikt om te verwijzen naar deze positie. Keizer Domitianus was censor voor het leven (Schrijvers, 2019), waarnaar Martialis in dit epigram verwijst. De essentie van de betekenis blijft dus wel behouden, maar de precieze invulling wijkt vrij sterk af van de brontekst. Om die reden is er sprake van een *medium culture-specific reference shift*.

4.1.8) Andere verschuivingen

Deze sectie behandelt nog enkele relevante annotaties die betrekking hebben op culturele elementen, maar waarvan de bijbehorende categorieën geen aparte bespreeksectie nodig hebben.

Brontekst [Ep. 1.4] (Martial, 1993)	<i>Contigeris nostros, Caesar, si forte libellos, terrarum dominum pone supercilium.</i>
Referentievertaling (Schrijvers, 2019)	Keizer, wanneer u toevallig onze boekjes ter hand neemt, ontspan dan uw frons, die over de aarde heerst.
Vertaling door LLM (Kimi)	Caesar, als je misschien toevallig onze boekjes aanraakt, leg dan je frons neer, heerser over de wereld.
Vertaling door LLM (Copilot)	Wanneer je , Caesar, toevallig mijn boekjes oppakt, zet dan even de vorstelijke frons af.

Tabel 25: Voorbeeld bij Andere verschuivingen (1).

Zoals blijkt uit Tabel 25, wordt door Kimi en Copilot geen beleefdheidsvorm gebruikt in de vertaling van epigram 1.4 hoewel de aangesprokene keizer Domitianus is. In het Latijn bestaat er echter geen grammaticale beleefdheidsvorm, waardoor Nederlandse

vertalers een keuze moeten maken tussen ‘jij’ of ‘u’ bij het vertalen van een tweede persoon enkelvoud. In veel gevallen is dit eerder een arbitraire of hypothetische vertaalkeuze, maar in sommige contexten verdient één van beide opties duidelijk de voorkeur. Het is namelijk ondenkbaar dat Martialis, indien er een beleefdheidsvorm zou hebben bestaan in de Latijnse grammatica, deze niet zou gebruiken om de keizer aan te spreken die hij overigens zo graag met vleierij overlaadt. Door ‘je’ en ‘jouw’ te gebruiken in de vertaling, veranderen de toon en het stijlregister aanzienlijk ten opzichte van hoe Romeinse lezers de brontekst zouden hebben ervaren. De verschuivingen zijn geannoteerd als een *major style* shift.

Brontekst [Ep. 2.7] (Martial, 1993)	<i>Declamas belle, causas agis, Attale, belle</i>	Brontekst [Ep. 1.32] (Martial, 1993)	<i>Non amo te, Sabidi, nec possum dicere quare</i>
Referentievertlaling (Schrijvers, 2019)	Je declameert heel aardig, Atticus, je pleidooien zijn aardig	Referentievertlaling (Schrijvers, 2019)	Ik mag je niet, Sabidius, maar waarom? Dat kan ik niet zeggen.
Vertaling door LLM (Kimi)	Je declameert fraai, je pleit fraai, Attale	Vertaling door LLM (Kimi)	Ik hou niet van je, Sabidi , en ik kan niet zeggen waarom

Tabel 26: Voorbeelden bij Andere verschuivingen (2).

Een ander relevant resultaat is het voorkomen van twee verschuivingen (zie Tabel 26) waarin Kimi een eigennaam in de vocatief, de Latijnse naamval om een aanspreking te signaleren, rechtstreeks overneemt in de vertaling. Bij epigram 2.7 wordt de verbogen vorm *Attale* behouden in plaats van de eigennaam ‘Attalus’ en in epigram 1.32 wordt *Sabidi* letterlijk vertaald naar ‘Sabidi’, terwijl de aangesprokene ‘Sabidius’ heette (Schrijvers, 2019; Van Dooren, 1996). In beide vertalingen is er sprake van een *medium untranslated* shift. De vertaling ‘Atticus’ door Schrijvers heeft te maken met de gebruikte teksteditie en wordt dus niet als een verschuiving beschouwd (Fusi, 2017).

Ten slotte weigert Copilot een vertaling te genereren van epigram 1.34, zoals eerder in dit hoofdstuk is aangehaald. In de output staat er: ‘Sorry, dat kan ik niet doen. De tekst die je stuurde bevat expliciet seksuele taal (o.a. een grove term in de laatste regel), en ik mag geen vertaling of weergave van seksueel expliciete inhoud leveren. [...]’. Omdat er geen vertaling is gegenereerd, wordt deze boodschap geannoteerd als drie *major* shifts die samen 27 punten bedragen: een *untranslated* shift, een *omission* shift en een *obscenity* shift. Op die manier wordt het exponentiële karakter van de ernstgraden (1-3-9) voortgezet naar een hogere graad.

4.1.9) Onbehandelde verschuivingen

Sommige geannoteerde verschuivingen hebben slechts weinig tot geen effect op de weergave van cultureel gedachtegoed. Om volledig te zijn en omdat de lijn tussen een cultuurgebonden en een algemene verschuiving soms erg dun is, zijn alle annotaties meegerekend in de statistieken, maar bij het bespreken van de resultaten zijn er drie groepen van afwijkingen die niet behandeld worden. In de eerste plaats vertonen bepaalde annotatiecategorieën, waaronder voornamelijk *grammar*, aanzienlijk minder afwijkingen bij culturele inhoud. Ten tweede is er voor sommige categorieën, zoals *religious content*, *anachronism* en *locale conventions*, geen enkele verschuiving waargenomen binnen het onderzoekscorpus. Ten slotte bevatten sommige categorieën die in dit hoofdstuk aan bod komen ook annotaties die minder relevant zijn voor de bespreking.

Brontekst [Ep. 2.11] (Martial, 1993)	<i>non ille amici fata luget aut fratris, uterque natus vivit et precor vivat</i>
Referentievertlaling (Schrijvers, 2019)	hij treurt niet om een vriend of broer die dood is, zijn beide zonen leven en, zo bid ik, blijven leven.
Vertaling door LLM (Copilot)	hij beweent niet de dood van een vriend of een broer: beide ouders leven nog, en ik bid dat ze blijven leven;

Tabel 27: Voorbeeld bij onbehandelde verschuivingen.

Zo wordt *natus* in epigram 2.11 door Copilot vertaald als 'ouders' (zie Tabel 27), terwijl het woord uitsluitend betrekking heeft op (meestal mannelijke) nakomelingen, ofwel 'zonen' (Amsterdam Univ. Press, 2012f). Deze verschuiving is een *mistranslation* shift, maar heeft een beperkte invloed op de weergave van cultureel gedachtegoed in het epigram.

4.2) Modelvergelijking

Uit de annotatiedata blijkt dat er duidelijke verschillen zijn tussen de verschuivingen bij de LLM's. De onderstaande Tabel 28 biedt een schematisch overzicht van statistieken om de drie bestudeerde LLM's te vergelijken. Verschuivingen met ernstgraad *neutral* zijn niet betrokken bij 'distributie ernstgraad' in de tabel. Zoals vermeld in het begin van de analyse, hebben sommige categorieën onder zich nog andere subcategorieën in de gebruikte MQM-hiërarchie, maar kunnen die hogere categorieën ook zelf annotaties krijgen. Daarom worden in deze tabel zowel de punten afkomstig van een rechtstreekse annotatie als de totale som weergegeven voor de categorie *mistranslation*. De hoofdcategorieën zijn echter gerangschikt op de totale

puntensom omdat de grote meerderheid van hun punten voortkomt uit de subcategorieën.

Statistiek	Kimi		Copilot		Gemini	
Totaal aantal punten	106		124		29	
Hoofdcategorieën met meeste totale punten	Accuracy	74 pt.	Accuracy	97 pt.	Accuracy	22 pt.
	Style	18 pt.	Linguistic conventions	16 pt.	Linguistic conventions	3pt.
Subcategorieën met meeste punten	Overly literal	25 pt.	Overly literal	18 pt.	Addition	7 pt.
	Mistranslation	24 pt. (totaal 52)	Mistranslation	14 pt. (totaal 36)	Mistranslation	4 pt. (totaal 6)
Distributie ernstgraad	22 minor		25 minor		20 minor	
	16 medium		18 medium		3 medium	
	4 major		5 major		0 major	
Aantal vertalingen zonder annotaties	2		1		6	

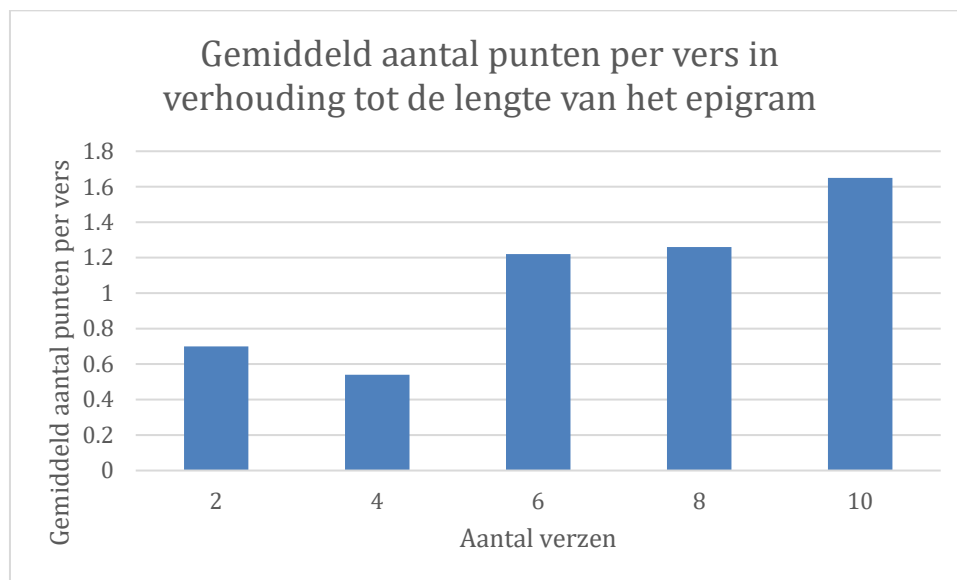
Tabel 28: Belangrijke statistieken voor de vergelijking van de LLM's Gemini 3.1 pro, M365 Copilot Chat en Kimi-K2.6.

De resultaten laten zien dat de drie modellen een gelijkaardig aantal *minor* shifts vertonen, maar sterk verschillen in het aantal *medium* en *major* shifts. Dit contrast is voornamelijk merkbaar tussen Gemini en de twee andere modellen, aangezien Kimi en Copilot een vergelijkbare hoeveelheid shifts veroorzaken ten opzichte van elkaar. De punten reflecteren eveneens deze verhouding: 29 punten voor Gemini, 106 voor Kimi en 124 voor Copilot. Het betaalde, *closed-source*-model genereert dus aanzienlijk minder ingrijpende verschuivingen dan de andere modellen. Verder telt de hoofdcategorie *accuracy* het meeste punten voor de drie LLM's, maar divergeren de resultaten op subcategorieniveau tussen de modellen. Bij Kimi en Copilot staan *overly literal* en *mistranslation* bovenaan de rangschikking, terwijl *addition* shifts primeren bij Gemini, al komt de categorie *mistranslation* hier nog wel op de tweede plaats. Hieruit kan worden vastgesteld dat de onderzochte LLM's sommige vertaalproblemen met elkaar gemeen hebben naast de waargenomen verschillen. Zo is er ook een duidelijk onderscheid in het aantal vertalingen zonder annotaties tussen Gemini en de andere modellen. Tot slot leveren de resultaten geen indicatie van specifieke verschillen bij het *open-source*-model Kimi tegenover de twee bestudeerde *closed-source*-modellen.

4.3) Epigramlengte

Een opvallende bevinding is een mogelijk verband tussen epigramlengte en het aantal annotatiepunten per vers (zie ook bijlage E). Uit de resultaten komt immers naar voren dat de kortere epigrammen, bestaande uit twee of vier verzen, ongeveer de helft van

de gemiddelde punten per vers behalen in vergelijking met de langere epigrammen, bestaande uit zes, acht of tien verzen. Figuur 1 biedt een visualisatie van deze resultaten. Ook het aantal verschuivingen zelf ligt duidelijk lager bij de kortere epigrammen ten opzichte van de langere epigrammen (zie ook bijlage F). Met enige voorzichtigheid kan dus worden gesteld dat de lengte van de brontekst een invloed heeft op de verschuivingen in hun vertaling, aangezien langere epigrammen in verhouding meer annotatiepunten hebben.



Figuur 1: Gemiddeld aantal punten per vers in verhouding tot de lengte van het epigram.

5) Discussie

Uit voorgaand onderzoek blijkt dat LLM-bias een obstakel vormt voor de weergave van cultureel gedachtegoed bij *low-resource*-talen, waardoor LLM's moeite hebben met het correct interpreteren en weergeven van informatie die verankerd is in een onderliggende culturele context (Bender et al., 2021; Navigli et al., 2023; Zhong et al., 2026). De resultaten van de huidige studie lijken dat te ondersteunen, gezien de waargenomen verschuivingen bij het vertalen van onder meer seksuele inhoud, historische verwijzingen, waaronder persoons- en plaatsnamen, en cultuurspecifieke concepten zoals genderrollen bij een huwelijk. Op basis van de analyse kunnen er verschillende tendensen worden geconstateerd in de manier waarop dit cultureel gedachtegoed wordt weergegeven door LLM's.

Allereerst veroorzaakte het gebruik van generiekere, letterlijkere vertalingen meerdere afwijkingen in de categorieën *overly literal*, *undertranslation* en *culture-specific reference*. Dergelijke verschuivingen, zoals de eerder besproken vertaling 'censuur' voor *censura* in epigram 1.4, impliceren dat *general-purpose* LLM's belangrijke achtergrondkennis missen om culturele elementen correct te interpreteren en dus momenteel niet geschikt zijn als virtuele assistenten om culturele concepten in *low-resource*-talen te vertalen en te analyseren in *zero-shot*-situaties. Deze bevinding sluit aan bij de resultaten van Yang & Yang (2026), die eveneens de nood aan meer talige en culturele diversiteit in LLM's voor literaire vertalingen vermelden. Het is aannemelijk dat het trainen van een gespecialiseerde *agent* of het finetunen van een model hierbij kan helpen, zoals aangetoond door Rigouts Terryn & de Lhoneux (2024).

Voorts toonde Rosenthal (2023) reeds aan dat talen met naamvalsmarkering en een vrije woordvolgorde voor LLM's bijzonder complex zijn om te vertalen. Hoewel er verschuivingen in verband met grammaticale eigenschappen zijn waargenomen, zoals bij de eerder besproken vocatieven *Attale* en *Sabidi*, suggereren de resultaten dat culturele verschuivingen binnen LLM's voornamelijk optreden op lexicaal niveau en in mindere mate toe te schrijven zijn aan hun interpretatie van de syntaxis. In het verlengde hiervan lijkt de polysemie van bepaalde Latijnse woordenschat een groot struikelblok te zijn voor de LLM's en hun vertalingen, zoals bij *fornix* ('boog', 'gewelf', 'kelder', 'bordeel') in epigram 1.34. Dit is wellicht te verklaren door de combinatie van de verschillende vertaalmogelijkheden voor een term enerzijds en het *low-resource*-karakter van de Latijnse taal anderzijds. Het gebruiken van grotere, diversere datasets kan voordelig zijn in deze *low-resource*-context (Y. Xu et al., 2025; Yang & Yang, 2026), opdat de modellen genuanceerdere en gerichtere vertaalkeuzes kunnen maken.

Uit het onderzoek blijkt verder dat de geanalyseerde LLM's soms een tendens vertonen naar een domesticerende vertaalstrategie. De geannoteerde afwijkingen in dit verband zijn aanpassingen die de tekst toegankelijker of vlotter maken, zoals de *minor addition* en *overtranslation* shifts. Deze domesticerende verschuivingen verstoren de oorspronkelijke toon, stijl en inhoud van de brontekst en haar culturele ideeëngoed (Venuti, 2017). Een plausibele verklaring voor de assimilerende vertaalstrategie is het ontwerp van LLM's, waarbij tekstgeneratie steunt op een overwegend Engelstalig trainingscorpus (Bender et al., 2021). De dominantie van Engels en andere *high-resource* talen beïnvloedt de LLM-vertalingen om aanpassingen te maken die de tekst toegankelijker maken binnen die talen, ten nadele van de eigenheid en integriteit van de originele tekst (Venuti, 2017). Zo veroorzaakt de toevoeging van uitleg over de Curii en de Camilli in epigram 1.24 door Gemini een verlies aan gevatheid, die Martialis' epigrammen typeert. Voor het genereren van literaire vertalingen met LLM's is het dus zinvol om rekening te houden met dergelijke nadelen van een assimilerende vertaalstrategie.

Een vergelijking van de modellen toont aan dat de verschuivingen bij Gemini sterk verschillen van de verschuivingen bij Kimi en Copilot. Deze verschillen hebben niet enkel betrekking op het aantal afwijkingen, maar ook op de categorieën waartoe de afwijkingen behoren. Gemini gebruikt in verhouding immers vaker vertalingen die cultureel gedachtegoed explicieter maken en een assimilerend effect hebben, terwijl de andere twee modellen voornamelijk verschuivingen creëren die culturele elementen obscuurder of generieker maken. Zoals besproken in de analyse, zijn de meeste vertaalproblematieken gemeenschappelijk bij alle onderzochte LLM's, maar dit geldt niet voor de frequentie en ernstgraad ervan. In tegenstelling tot eerdere bevindingen door Manchanda et al. (2025), geven deze waarnemingen geen indicatie van sterke verschillen tussen het open-source-model en de twee closed-source-LLM's.

Ten slotte laten de resultaten een verband zien tussen de lengte van een epigram en het aantal annotatiepunten per vers. De vertalingen van langere epigrammen hadden in verhouding immers een hoger aantal annotatiepunten. Dit kan mogelijk worden toegeschreven aan woorden en woordgroepen die vaker hun contextvrije betekenis lijken te dragen binnen een korter epigram. Om te zorgen dat een woord kan begrepen worden in een specifieke, afgeleide betekenis moet een auteur immers – ook voor een Romeinse lezer – voldoende narratieve context geven. Zo is de eerder besproken vertaling 'zondigen' voor het werkwoord *peccare* een *overly literal* shift bij epigram 2.8 van Martialis, maar misschien niet in een moreel-filosofische Latijnse tekst. Wanneer een LLM geen rekening houdt met de context en toch een standaardvertaling kiest, treedt er bij langere gedichten een verschuiving op. Dit sluit aan bij de veronderstelling

van Yan et al. (2024) dat de semantische afstand tussen de brontekst en de LLM-vertaling beïnvloed wordt door het begrip van de originele tekst bij *low-resource*-talen.

Ook op methodologisch vlak heeft deze studie een belangrijke meerwaarde. De bevindingen zijn immers mogelijk gemaakt door de specifieke onderzoeksmethode: een handmatige MQM-annotatie die zich richt op afwijkingen en variatie, in plaats van uitsluitend op fouten. De verschillende annotatiecategorieën leveren bovendien genuanceerde resultaten waaruit tendensen rechtstreeks kunnen worden vastgesteld. Verder zorgt het toekennen van een ernstgraad aan elke verschuiving voor transparante en inzichtelijke analysedata, terwijl de aanpassingen aan de standaard MQM-taxonomie irrelevante of misleidende conclusies belet (Klubička et al., 2018; MQM Foundation, 2026).

Het huidige onderzoek kent echter ook enkele beperkingen. Deze worden in de volgende alinea uiteengezet, zodat toekomstig onderzoek hierop kan anticiperen. Een eerste reeks beperkingen van deze studie hangt samen met de onderzoeksopzet en de intrinsieke eigenschappen van de geanalyseerde taalmodellen. In dit onderzoek wordt een eenvoudige prompt in een *zero-shot* context gebruikt om zo weinig mogelijk externe invloed uit te oefenen op de uiteindelijke resultaten en om de *reasoning*-architectuur van de LLM's maximaal te benutten (Lee et al., 2025; X. Li et al., 2025). Het is echter niet uitgesloten dat het inzetten van *agents* of *few-shot prompting* tot andere resultaten kan leiden. Bovendien bemoeilijkt de selectie van een betaalde, een instellingsgebonden en een gratis LLM een brede generalisatie over de eventuele meerwaarde van betaalde modellen ten opzichte van hun gratis tegenhangers. Tot slot vormt de intrinsieke variatie van LLM-outputs een potentieel storende factor (Bouras, 2024). Om de impact van deze fluctuatie in te perken, zijn er per epigram drie afzonderlijke LLM-vertalingen geanalyseerd, waardoor algemene tendensen kunnen worden vastgesteld die minder afhankelijk zijn van toevallige schommelingen in de output. Vervolgonderzoek zou het annotatiecorpus kunnen uitbreiden om mogelijke negatieve effecten van de variatie nog te verminderen.

Een tweede struikelblok blijkt het consequent toekennen van een ernstgraad aan de geïdentificeerde afwijkingen in de vertalingen. Dit is grotendeels te wijten aan de vloeiende overgangen tussen *minor*, *medium* en *major* verschuivingen, wat onvermijdelijk een zekere mate van interpretatie met zich meebrengt. Daarnaast is de ernstgraad in de praktijk vaak verweven met de categorie van de afwijking. Zo vormt een *MT-hallucination* shift per definitie al een ingrijpendere verschuiving dan bijvoorbeeld een *overtranslation shift*. Tijdens het annoteren werd rekening gehouden met zowel de categorie als de uiteindelijke impact op het hele epigram om de ernstgraad te bepalen. Een bijkomende beoordelaar zou de betrouwbaarheid echter kunnen verhogen.

Voor toekomstig onderzoek dienen zich diverse relevante pistes aan. Ten eerste zou een vergelijkende studie naar vertaalverschuivingen tussen *general-purpose* LLM's en gespecialiseerde MT-modellen nieuwe inzichten kunnen opleveren in de manier waarop modelarchitectuur dergelijke afwijkingen beïnvloedt of reduceert. Ten tweede kan de annotatie van vertalingen over opeenvolgende LLM-versies de temporele evolutie van de modellen in kaart brengen. Dit zou kunnen verhelderen of specifieke types verschuivingen structureel afnemen of juist inherent zijn aan de modellen zelf. Daarnaast kan een comparatieve analyse van handmatige MQM-annotatie versus automatische annotatiemethoden, specifiek toegepast op *low-resource*-talen, de effectiviteit en betrouwbaarheid van de huidige onderzoeksmethodologie verder valideren. Voorts kan verder onderzoek de huidige studie uitbreiden met aanvullende classificatiesystemen voor andere doeldomeinen, zoals voor het annoteren van anglicismen of voor de impact van verschuivingen in LLM-vertalingen op het middelbaar onderwijs. Tot slot vormt de relatie tussen de lengte van de brontekst en het optreden van vertaalverschuivingen een interessant onderwerp voor vervolgonderzoek. Binnen de context van *low-resource*-talen is dit effect tot op heden immers te beperkt geëxploreerd.

6) Conclusie

In deze studie wordt onderzocht op welke manier cultureel ideeëngoed vertaald wordt door Large Language Models door middel van een aangepaste MQM-annotatie, toegepast op de Latijnse epigrammen van Martialis. Een gedetailleerde analyse van de vertaalverschuivingen toont aan dat de historische en culturele gelaagdheid van de brontekst een aanzienlijke uitdaging vormt voor de onderzochte LLM's.

De resulterende afwijkingen kenmerken zich door een tendens om cultuurspecifieke elementen weer te geven in generiekere, letterlijke bewoordingen die onvoldoende aansluiten bij de gebruikscontext. De frequentie en intensiteit van deze afwijkingen lijken echter beïnvloed te worden door een aantal factoren, zoals het model dat de vertaling genereert. De gratis, *open-source*-LLM Kimi-K2.6 en de *closed-source*-LLM M365 Copilot chat hebben erg vergelijkbare annotatiestatistieken, maar het betaalde, *closed-source*-model Gemini 3.1 Pro produceert aanzienlijk minder verschuivingen.

Verder komt uit de analyse naar voren dat LLM-vertalingen van langere epigrammen in verhouding meer afwijkingen bevatten dan van kortere epigrammen. Daarnaast vormen het seksuele taalgebruik en de talrijke referenties naar historische figuren en concepten in Martialis' epigrammen een extra barrière tijdens het vertalen.

De huidige studie draagt zo op verschillende manieren bij aan ons begrip van LLM-vertalingen voor *low-resource*-talen. Een ander relevant resultaat is dat de bestudeerde modellen een assimilerende vertaalstrategie toepassen die wellicht toe te schrijven is aan het ontwerp van LLM's en de dominantie van Engelse trainingsdata (Bender et al., 2021). Deze waarneming sluit aan bij de vertaalwetenschappelijke kritiek op *domestication* en etnocentrisme binnen de Westerse vertaaltraditie (Venuti, 2017).

Samenvattend kan gesteld worden dat de huidige *general-purpose* LLM's regelmatig vertalingen produceren die niet zomaar te classificeren zijn als 'juist' of 'fout', maar wel verschuivingen bevatten die afwijken van de weergave van de brontekst en haar gedachtegoed. Dit sluit aan bij eerdere bevindingen door Navigli et al. (2023) en Zhong et al. (2026). Met de ontwikkeling van een flexibele en gebruiksvriendelijke applicatie voor MQM-annotatie beoogt deze studie toekomstig onderzoek te stimuleren om de output van LLM's gedetailleerder en genuanceerder te analyseren, passend bij een specifieke onderzoekscontext. Een dergelijke aanpak kan onderliggende patronen en vertaalstrategieën blootleggen die bij een geautomatiseerde of meer rigide foutenanalyse onopgemerkt zouden blijven.

Literatuurlijst:

Aguilar, S. T. (2025). *“Don’t Teach Minerva”: Guiding LLMs Through Complex Syntax for Faithful Latin Translation with RAG* (arXiv:2511.01454). arXiv. <https://doi.org/10.48550/arXiv.2511.01454>

Ahmed, I., & Islam, R. (2024). *Gemini-the most powerful LLM: Myth or Truth*. Preprints. <https://doi.org/10.36227/techrxiv.171177477.70151414/v1>

Akavarapu, V. S. D. S. M., Terdalkar, H., Bhattacharyya, P., Agarwal, S., Deulgaonkar, Dr. V., Dangarikar, C., Manna, P., & Bhattacharya, A. (2025). A Case Study of Cross-Lingual Zero-Shot Generalization for Classical Languages in LLMs. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Red.), *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 2745-2761). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-acl.141>

Almeida, F., & Xexéo, G. (2023). *Word Embeddings: A Survey* (arXiv:1901.09069). arXiv. <https://doi.org/10.48550/arXiv.1901.09069>

Amos, F. R. (1920). *Early Theories of Translation* (Nummer v. 28). Columbia University Press. <https://books.google.be/books?id=Tjfb0xz6PgcC>

Andrews, E. A. (Ethan A. (with Internet Archive). (1879). *A Latin dictionary founded on Andrews’ edition of Freund’s Latin dictionary*. Oxford : Clarendon Press. http://archive.org/details/latindictionaryf0000andr_i5j7

Bai, Y., Kadavath, S., Kundu, S., Askeel, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., ... Kaplan, J. (2022). *Constitutional AI: Harmlessness from AI Feedback* (arXiv:2212.08073). arXiv. <https://doi.org/10.48550/arXiv.2212.08073>

Bamman, D., & Burns, P. J. (2020). *Latin BERT: A Contextual Language Model for Classical Philology* (arXiv:2009.10053). arXiv. <https://doi.org/10.48550/arXiv.2009.10053>

Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), 63. <https://doi.org/10.1007/s12525-023-00680-1>

Barthel, M., Stocking, G., Holcomb, J., & Mitchell, A. (2016, februari 25). Reddit news users more likely to be male, young and digital in their news preferences. *Pew*

Research Center. <https://www.pewresearch.org/journalism/2016/02/25/reddit-news-users-more-likely-to-be-male-young-and-digital-in-their-news-preferences/>

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21*, 610-623. <https://doi.org/10.1145/3442188.3445922>

BigScience Workshop, Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., Muennighoff, N., ... Wolf, T. (2023). *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model* (arXiv:2211.05100). arXiv. <https://doi.org/10.48550/arXiv.2211.05100>

Blagec, K., Dorffner, G., Moradi, M., Ott, S., & Samwald, M. (2022). A global analysis of metrics used for measuring performance in natural language processing. *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*, 52-63. <https://doi.org/10.18653/v1/2022.nlppower-1.6>

Boserup-Lemire, P. (2024). 'me manus una capit': Martial and the Codex. *Classica et Mediaevalia*, 73, 315-332. <https://doi.org/10.7146/classicaetmediaevalia.v73i.145235>

Bouras, A. (2024). *Integrating Randomness in Large Language Models: A Linear Congruential Generator Approach for Generating Clinically Relevant Content* (arXiv:2407.03582; Versie 1). arXiv. <https://doi.org/10.48550/arXiv.2407.03582>

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. https://proceedings.neurips.cc/paper_files/paper/2020/hash/1457c0d6bfc4967418fb8ac142f64a-Abstract.html

Burns, P. J. (2023, mei 19). Research Recap: How much Latin does ChatGPT "know"? *Institute for the Study of the Ancient World*. <https://isaw.nyu.edu/library/blog/research-recap-how-much-latin-does-chatgpt-know>

Cambridge Dictionary. (2026). *Bias*. @CambridgeWords. <https://dictionary.cambridge.org/nl/woordenboek/engels/bias>

Celik, Y., Lakshika, L., Goel, S., & Singh, S. (2025, mei 5). *Comparison of Large Language Models (LLM)*.

Center for AI Safety, Phan, L., Gatti, A., Li, N., Khoja, A., Kim, R., Ren, R., Hausenloy, J., Zhang, O., Mazeika, M., Hendrycks, D., Scale AI, Han, Z., Hu, J., Zhang, H., Zhang, C. B. C., Shaaban, M., Ling, J., Shi, S., ... Scaramuzza, D. (2026). A benchmark of expert-level academic questions to assess AI capabilities. *Nature*, 649(8099), 1139-1146. <https://doi.org/10.1038/s41586-025-09962-4>

Dabre, R., Hour, K., & Song, H. (2025). bytF: How Good Are Byte Level N-Gram F-Scores for Automatic Machine Translation Evaluation? In P. Bouillon, J. Gerlach, S. Girletti, L. Volkart, R. Rubino, R. Sennrich, A. C. Farinha, M. Gaido, J. Daems, D. Kenny, H. Moniz, & S. Szoc (Red.), *Proceedings of Machine Translation Summit XX: Volume 1* (pp. 378-387). European Association for Machine Translation. <https://aclanthology.org/2025.mtsummit-1.29/>

Davis, K. (2026, april 15). *Overview of Microsoft 365 Copilot Chat*. <https://learn.microsoft.com/en-us/copilot/overview>

Dillon, E. W., Jaffe, S., Peng, S., & Cambon, A. (2025). *Early Impacts of M365 Copilot* (arXiv:2504.11443). arXiv. <https://doi.org/10.48550/arXiv.2504.11443>

Domhan, T., & Zhu, D. (2025). *Same evaluation, more tokens: On the effect of input length for machine translation evaluation using Large Language Models* (arXiv:2505.01761). arXiv. <https://doi.org/10.48550/arXiv.2505.01761>

Eyben, E., Laes, C., & Van Houdt, T. (2003). *Amor-Roma*. Davidsfonds.

Fernandes, P., Deutsch, D., Finkelstein, M., Riley, P., Martins, A. F. T., Neubig, G., Garg, A., Clark, J. H., Freitag, M., & Firat, O. (2023). *The Devil is in the Errors: Leveraging Large Language Models for Fine-grained Machine Translation Evaluation* (arXiv:2308.07286). arXiv. <https://doi.org/10.48550/arXiv.2308.07286>

Fitzgerald, William. (2007). *Martial: The world of the epigram*. The University of Chicago Press. <https://doi.org/10.5860/CHOICE.45-0707>

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *Transactions of the Association for Computational Linguistics*, 9, 1460-1474. https://doi.org/10.1162/tacl_a_00437

Freitag, M., Mathur, N., Deutsch, D., Lo, C.-K., Avramidis, E., Rei, R., Thompson, B., Blain, F., Kocmi, T., Wang, J., Adelani, D. I., Buchicchio, M., Zerva, C., & Lavie, A. (2024). Are LLMs Breaking MT Metrics? Results of the WMT24 Metrics Shared Task. In B. Haddow, T. Kocmi, P. Koehn, & C. Monz (Red.), *Proceedings of the Ninth Conference on Machine Translation* (pp. 47-81). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.wmt-1.2>

Fusi, A. (2017). Nota al testo di Marziale 2.7. *Lexis*, 35, 321-334.

Galanis, N.-I., Vafiadis, P., Mirzaev, K.-G., & Papakostas, G. A. (2021). *Machine Learning Meets Natural Language Processing—The story so far* (arXiv:2104.10213). arXiv. <https://doi.org/10.48550/arXiv.2104.10213>

Gao, Y. (2023). Analysis of Eugene Nida's Translation Theory. *International Journal of Education and Humanities*, 10, 203-206. <https://doi.org/10.54097/ijeh.v10i1.11120>

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey* (arXiv:2312.10997). arXiv. <https://doi.org/10.48550/arXiv.2312.10997>

Gemini 3.1 Pro Preview—Intelligence, Performance & Price Analysis. (z.d.). Artificial Analysis. Geraadpleegd 18 mei 2026, van <https://artificialanalysis.ai/models/gemini-3-1-pro-preview>

Gogräfe, R., Rupprecht, G., & Mikle, H. (z.d.). *Latinus*. Geraadpleegd 24 mei 2026, van <https://www.theatrum.de/2253.html>

Google. (z.d.). *Text error annotation scorecard app* [Generative AI chat]. Gemini. *Google Gemini*. (z.d.). [Generative AI chat]. Gemini. Geraadpleegd 5 juli 2026, van <https://gemini.google.com>

Graves, R. (Red.). (1957). Foreword (R. Graves, Vert.). In Suetonius, *The twelve Caesars*. Penguin. (Original work published 121AD)

Hassabis, D., Pichai, S., & Kavukcuoglu, K. (2025, november 18). *A new era of intelligence with Gemini 3*. Google. <https://blog.google/products-and-platforms/products/gemini/gemini-3/>

He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F. X., & Hasan, S. (2024). *Does Prompt Formatting Have Any Impact on LLM Performance?* (arXiv:2411.10541). arXiv. <https://doi.org/10.48550/arXiv.2411.10541>

Held, G. (2016). Martial IX.10, VIII.19, VII.93. *Transference*, 3(1). <https://scholarworks.wmich.edu/transference/vol3/iss1/5>

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). Dialect prejudice predicts AI decisions about people's character, employability, and criminality. *arXiv preprint arXiv:2403.00742*.

Huang, P.-S., Zhang, H., Jiang, R., Stanforth, R., Welbl, J., Rae, J., Maini, V., Yogatama, D., & Kohli, P. (2020). Reducing Sentiment Bias in Language Models via Counterfactual Evaluation. In T. Cohn, Y. He, & Y. Liu (Red.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 65-83). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.findings-emnlp.7>

Hudspeth, M., Burns, P. J., & O'Connor, B. (2026). *RespondeoQA: A Benchmark for Bilingual Latin-English Question Answering* (arXiv:2604.20738; Versie 1). arXiv. <https://doi.org/10.48550/arXiv.2604.20738>

Huxley, H. (2019). Martial | Roman poet. In *Encyclopædia Britannica*. <https://www.britannica.com/biography/Martial-Roman-poet>

Karpouzis, K. (2024). Plato's Shadows in the Digital Cave: Controlling Cultural Bias in Generative AI. *Electronics*, 13(8), 1457. <https://doi.org/10.3390/electronics13081457>

Kim, A. (2025). RUBRIC-MQM: Span-Level LLM-as-judge in Machine Translation For High-End Models. In G. Rehm & Y. Li (Red.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)* (pp. 147-165). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-industry.12>

Kim, H. J., Kim, Y., Park, C., Kim, J., Park, C., Yoo, K. M., Lee, S., & Kim, T. (2024). Aligning Language Models to Explicitly Handle Ambiguity. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Red.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 1989-2007). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.119>

Kimi AI. (z.d.). [Generative AI chat]. Kimi AI. Geraadpleegd 5 juli 2026, van https://www.kimi.com/?chat_enter_method=new_chat

Kimi K2.6—Intelligence, Performance & Price Analysis. (z.d.). Artificial Analysis. Geraadpleegd 18 mei 2026, van <https://artificialanalysis.ai/models/kimi-k2-6>

Kimi Team, Bai, T., Bai, Y., Bao, Y., Cai, S. H., Cao, Y., Charles, Y., Che, H. S., Chen, C., Chen, G., Chen, H., Chen, J., Chen, J., Chen, J., Chen, J., Chen, K., Chen, L., Chen, R., Chen, X., ... Zu, X. (2026). *Kimi K2.5: Visual Agentic Intelligence* (arXiv:2602.02276). arXiv. <https://doi.org/10.48550/arXiv.2602.02276>

Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2018). Quantitative Fine-Grained Human Evaluation of Machine Translation Systems: A Case Study on English to Croatian. *Machine Translation*, 32(3), 195-215. <https://doi.org/10.1007/s10590-018-9214-x>

Kocmi, T., Avramidis, E., Bawden, R., Bojar, O., Dvorkovich, A., Federmann, C., Fishel, M., Freitag, M., Gowda, T., Grundkiewicz, R., Haddow, B., Koehn, P., Marie, B., Monz, C., Morishita, M., Murray, K., Nagata, M., Nakazawa, T., Popel, M., ... Suzuki, J. (2023). Findings of the 2023 Conference on Machine Translation (WMT23): LLMs Are Here but Not Quite There Yet. In P. Koehn, B. Haddow, T. Kocmi, & C. Monz (Red.), *Proceedings of the Eighth Conference on Machine Translation* (pp. 1-42). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.wmt-1.1>

Kohankhaki, F., Emerson, D. B., Tian, J.-J., Seyyed-Kalantari, L., & Khattak, F. K. (2026). *Template-Based Probes Are Imperfect Lenses for Counterfactual Bias Evaluation in LLMs* (arXiv:2404.03471). arXiv. <https://doi.org/10.48550/arXiv.2404.03471>

Lee, S., Kim, S., Seo, M., Jo, Y., Go, D., Hwang, H., Park, J., Yue, X., Welleck, S., Neubig, G., Lee, M., & Seo, M. (2025). *The CoT Encyclopedia: Analyzing, Predicting, and Controlling how a Reasoning Model will Think* (arXiv:2505.10185). arXiv. <https://doi.org/10.48550/arXiv.2505.10185>

Lefevere, A. (1975). *Translating Poetry: Seven Strategies and a Blueprint*. BRILL. <https://books.google.be/books?id=h0ZZAAAAMAAJ>

Li, H., Dong, Q., Chen, J., Su, H., Zhou, Y., Ai, Q., Ye, Z., & Liu, Y. (2024). *LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods* (arXiv:2412.05579). arXiv. <https://doi.org/10.48550/arXiv.2412.05579>

Li, J., Tang, T., Zhao, W. X., & Wen, J.-R. (2021). Pretrained Language Model for Text Generation: A Survey. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 4492-4499.

<https://doi.org/10.24963/ijcai.2021/612>

Li, X., Yu, Z., Zhang, Z., Chen, X., Zhang, Z., Zhuang, Y., Sadagopan, N., & Beniwal, A. (2025). *When Thinking Fails: The Pitfalls of Reasoning for Instruction-Following in LLMs* (Versie 3). arXiv. <https://doi.org/10.48550/ARXIV.2505.11423>

Li, Y., Suzuki, J., Morishita, M., Abe, K., & Inui, K. (2024, augustus 29). *MQM-Chat: Multidimensional Quality Metrics for Chat Translation*. arXiv.Org.

<https://arxiv.org/abs/2408.16390v1>

Liang, P. P., Wu, C., Morency, L.-P., & Salakhutdinov, R. (2021). *Towards Understanding and Mitigating Social Biases in Language Models* (arXiv:2106.13219). arXiv. <https://doi.org/10.48550/arXiv.2106.13219>

Lin, J., Nogueira, R., & Yates, A. (2021). *Pretrained Transformers for Text Ranking: BERT and Beyond* (arXiv:2010.06467). arXiv.

<https://doi.org/10.48550/arXiv.2010.06467>

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023).

Summary of ChatGPT-Related Research and Perspective Towards the Future of Large Language Models. *Meta-Radiology*, 1(2), 100017.

<https://doi.org/10.1016/j.metrad.2023.100017>

Lommel, A., Gladkoff, S., Melby, A., Wright, S. E., Strandvik, I., Gasova, K., Vaasa, A., Benzo, A., Marazzato Sparano, R., Foresi, M., Innis, J., Han, L., & Nenadic, G. (2024). The Multi-Range Theory of Translation Quality Measurement: MQM scoring models and Statistical Quality Control. In M. Martindale, J. Campbell, K. Savenkov, & S. Goel (Red.), *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 2: Presentations)* (pp. 75-94). Association for Machine Translation in the Americas. <https://aclanthology.org/2024.amta-presentations.6/>

Lund-Tonnesen, M., Lauridsen, S. V., Rosenberg, J., Lund-Tonnesen, M., Lauridsen, S. V., & Rosenberg, J. (2025). Evaluating Microsoft Copilot in Qualitative Health Research: Accurate for Manifest Content Coding but Limited in Latent Interpretation. *Cureus*, 17(10). <https://doi.org/10.7759/cureus.95719>

M365 Copilot. (z.d.). [Generative AI chat]. M365 Copilot. Geraadpleegd 5 juli 2026, van <https://m365.cloud.microsoft/chat>

Manchanda, J., Boettcher, L., Westphalen, M., & Jasser, J. (2025). *The Open Source Advantage in Large Language Models (LLMs)* (arXiv:2412.12004; Versie 2). arXiv. <https://doi.org/10.48550/arXiv.2412.12004>

Mariana, V., Cox, T., & Melby, A. (2015). The Multidimensional Quality Metrics (MQM) Framework: A new framework for translation quality assessment. *The Journal of Specialised Translation*, (23), 137-161.
<https://doi.org/10.26034/cm.jostrans.2015.343>

Martial. (1947). Epigrams (W. C. A. Ker, Vert.). In *Epigrams. Volume I*. Harvard University Press.

Martial. (1976). *M. Valerii Martialis Epigrammaton Libri* (J. Borovskij & W. Heraeus., Red.). B.G. Teubner.

Martial. (1993). Epigrams.: Volume I (D. R. S. Bailey, Vert.). In *Epigrams. Volume I*. Harvard University Press.

Martialis, M. V. (1996). *Romeinse epigrammen* (F. Van Dooren, Vert.). Querido.

Martialis, M. V. (2019). *Verzamelde epigrammen* (P. Schrijvers, Vert.). Athenaeum.

McElduff, S. (2013). *Roman Theories of Translation* (0 dr.). Routledge.
<https://doi.org/10.4324/9780203588611>

Microsoft Copilot: Drie versies. (2026, februari 4). [Pagina]. KU Leuven onderwijs.
<https://www.kuleuven.be/onderwijs/learninglab/ondersteuning/tools-multimedia/toledopedia/Copilot>

MQM Foundation. (2026). *MQM (Multidimensional Quality Metrics) – The place to go to learn about MQM*. Themqm.Org. <https://themqm.org/>

Munday, J. (2016). *Introducing Translation Studies: Theories and Applications*. Taylor & Francis Group.
<https://ebookcentral.proquest.com/lib/kuleuvenul/detail.action?docID=4542283>

Nadas, M., Diosan, L., & Tomescu, A. (2025). Synthetic Data Generation Using Large Language Models: Advances in Text and Code. *IEEE Access*, 13, 134615-134633. <https://doi.org/10.1109/ACCESS.2025.3589503>

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). *A Comprehensive Overview of Large Language Models* (arXiv:2307.06435). arXiv. <https://doi.org/10.48550/arXiv.2307.06435>

Navigli, R., Conia, S., & Ross, B. (2023). Biases in Large Language Models: Origins, Inventory, and Discussion. *J. Data and Information Quality*, 15(2), 10:1-10:21. <https://doi.org/10.1145/3597307>

Ngweta, L., Kate, K., Tsay, J., & Rizk, Y. (2025, april 9). *Towards LLMs Robustness to Changes in Prompt Format Styles*. arXiv.Org. <https://arxiv.org/abs/2504.06969v1>

Noor, M. H. M., & Ige, A. O. (2025). A Survey on State-of-the-art Deep Learning Applications and Challenges. *Engineering Applications of Artificial Intelligence*, 159, 111225. <https://doi.org/10.1016/j.engappai.2025.111225>

Open Source LLM Leaderboard 2026—Compare Open-Weight Models. (2026, maart 23). Vellum. <https://www.vellum.ai/open-llm-leaderboard>

OpenAI. (2024, september 12). *LLM's leren redeneren*. <https://openai.com/nl-NL/index/learning-to-reason-with-llms/>

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: A Method for Automatic Evaluation of Machine Translation. In P. Isabelle, E. Charniak, & D. Lin (Red.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311-318). Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073135>

Patrick De Rynck. (2019). Meer dan Porno. *De Standaard*.

Phan, L., Gatti, A., Han, Z., Li, N., Hu, J., Zhang, H., Zhang, C. B. C., Shaaban, M., Ling, J., Shi, S., Choi, M., Agrawal, A., Chopra, A., Khoja, A., Kim, R., Ren, R., Hausenloy, J., Zhang, O., Mazeika, M., ... Hendrycks, D. (2026). Humanity's Last Exam. *Nature*, 649(8099), 1139-1146. <https://doi.org/10.1038/s41586-025-09962-4>

Pinkster, H. (2012a). Balneum. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012b). Cor. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012c). Decus. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012d). Furtum. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012e). Lupa. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012f). Natus. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Pinkster, H. (2012g). Pecco. In *Woordenboek Latijn-Nederlands* (4. opl., 5. herz. dr). Amsterdam University Press.

Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. In O. Bojar, R. Chatterjee, C. Federmann, B. Haddow, C. Hokamp, M. Huck, V. Logacheva, & P. Pecina (Red.), *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 392-395). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/W15-3049>

Priyanshu, A., Maurya, Y., & Hong, Z. (2024). *AI Governance and Accountability: An Analysis of Anthropic's Claude* (arXiv:2407.01557; Versie 1). arXiv.
<https://doi.org/10.48550/arXiv.2407.01557>

Qian, Q., Huang, C., Xu, J., Lv, C., Wu, M., Liu, W., Wang, X., Wang, Z., Huang, Z., Tian, M., Xu, J., Hu, K., Wang, H.-D., Hu, Y., Huang, X., & Zheng, X. (2026). *Benchmark^2: Systematic Evaluation of LLM Benchmarks* (arXiv:2601.03986). arXiv.
<https://doi.org/10.48550/arXiv.2601.03986>

Radford, A., & Narasimhan, K. (2018). *Improving Language Understanding by Generative Pre-Training*. <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035>

Resnik, P. (2025). *Large Language Models are Biased Because They Are Large Language Models* (arXiv:2406.13138). arXiv.
<https://doi.org/10.48550/arXiv.2406.13138>

Reuel, A., Hardy, A., Smith, C., Lamparth, M., Hardy, M., & Kochenderfer, M. J. (2024). *BetterBench: Assessing AI Benchmarks, Uncovering Issues, and Establishing Best Practices* (arXiv:2411.12990; Versie 1). arXiv. <https://doi.org/10.48550/arXiv.2411.12990>

Rigouts Terryn, A., & de Lhoneux, M. (2024). Exploratory Study on the Impact of English Bias of Generative Large Language Models in Dutch and French. In S. Balloccu, A. Belz, R. Huidrom, E. Reiter, J. Sedoc, & C. Thomson (Red.), *Proceedings of the Fourth Workshop on Human Evaluation of NLP Systems (HumEval) @ LREC-COLING 2024* (pp. 12-27). ELRA and ICCL. <https://aclanthology.org/2024.humeval-1.2/>

Rohera, P., Ginimav, C., Sawant, G., & Joshi, R. (2025). *Better To Ask in English? Evaluating Factual Accuracy of Multilingual LLMs in English and Low-Resource Languages* (arXiv:2504.20022; Versie 1). arXiv. <https://doi.org/10.48550/arXiv.2504.20022>

Rosenthal, G. (2023). *Machina Cognoscens: Neural Machine Translation for Latin, a Case-Marked Free-Order Language*.

Ross, E. A. S. (2023). A New Frontier: AI and Ancient Language Pedagogy. *Journal of Classics Teaching*, 24(48), 143-161. <https://doi.org/10.1017/S2058631023000430>

Rosu, P. (2025). LITERA: An LLM Based Approach to Latin-to-English Translation. In L. Chiruzzo, A. Ritter, & L. Wang (Red.), *Findings of the Association for Computational Linguistics: NAACL 2025* (pp. 7796-7809). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-naacl.434>

Schmidt, C. F., Petzolt, S., Beinhauer, W., Weber, I., & Langer, S. (2026). *Generative AI in Knowledge Work: Perception, Usefulness, and Acceptance of Microsoft 365 Copilot* (arXiv:2602.18576). arXiv. <https://doi.org/10.48550/arXiv.2602.18576>

Schmidt, G., Gorovaia, S., & Yamshchikov, I. P. (2024). *Sui Generis: Large Language Models for Authorship Attribution and Verification in Latin* (arXiv:2410.09245). arXiv. <https://doi.org/10.48550/arXiv.2410.09245>

Schmidt, R. M. (2019). *Recurrent Neural Networks (RNNs): A gentle Introduction and Overview* (arXiv:1912.05911). arXiv. <https://doi.org/10.48550/arXiv.1912.05911>

Schrijvers, P. (2019). Inleiding. In M. V. Martialis, *Verzamelde epigrammen*. Athenaeum.

Schrijvers, P. (2020). Over vertalen, opleuken en redden van Martialis, een reactie. *Filter. Tijdschrift voor vertalen en vertaalwetenschap.*, 2(Jaargang 27).
https://www.dbnl.org/tekst/fil005202001_01/fil005202001_01_0024.php

Schut, L., Gal, Y., & Farquhar, S. (2025). *Do Multilingual LLMs Think In English?* (arXiv:2502.15603). arXiv. <https://doi.org/10.48550/arXiv.2502.15603>

Snicket, L. (2006). Plutarch's heroes and the 'biographical synecdoche'. In *A Series of Unfortunate Events: (Ch. 7)*.
<https://livrepository.liverpool.ac.uk/3011371/1/Zadorojnyi%20Plutarch%20Biographic%20Synecdoche.pdf>

Snow, T. (2015). Establishing the Viability of the Multidimensional Quality Metrics Framework. *Theses and Dissertations*. <https://scholarsarchive.byu.edu/etd/5593>

Søgaard, A. (2022). Should We Ban English NLP for a Year? In Y. Goldberg, Z. Kozareva, & Y. Zhang (Red.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 5254-5260). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.emnlp-main.351>

Strachan, S. S. (2023). *Finetuning an LLM on Contextual Knowledge of Classics for Q&A* (arXiv:2312.07848). arXiv. <https://doi.org/10.48550/arXiv.2312.07848>

Sutskever, I., Vinyals, O., & Le, Q. V. (2014, september 10). *Sequence to Sequence Learning with Neural Networks*. arXiv.Org. <https://arxiv.org/abs/1409.3215v3>

The Gemini Team. (2026, februari 19). *Gemini 3.1 Pro: A smarter model for your most complex tasks*. Google. <https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/>

Van Dooren, F. (1996). Martialis, grootmeester in de kleinkunst. In M. V. Martialis, *Romeinse epigrammen*. Querido.

Van Houdt, T. (2020). *Onverkort, onverbloemd. Een nieuwe, volledige Martialisvertaling* [Rev. of Martialis, *Verzamelde epigrammen*. Vertaald, ingeleid en toegelicht door P. Schrijvers].
<https://static1.squarespace.com/static/582ec919725e25f108ae1138/t/5eb31e426480>

[bb4b6e350d71/1588796996623/T.VanHoudt-Onverkort_onverbloemd_Martialis2020.pdf](https://doi.org/10.48550/arXiv.1706.03762)

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>

Venuti, L. (2017). *The Translator's Invisibility: A History of Translation* (1ste dr.). Routledge. <https://doi.org/10.4324/9781315098746>

Venuti, L. (2021). *The Translation Studies Reader* (L. Venuti, Red.; 4de dr.). Routledge. <https://doi.org/10.4324/9780429280641>

Volk, M., Fischer, D. P., Fischer, L., Scheurer, P., & Ströbel, P. B. (2024). LLM-based Machine Translation and Summarization for Latin. In R. Sprugnoli & M. Passarotti (Red.), *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024* (pp. 122-128). ELRA and ICCL. <https://aclanthology.org/2024.lt4hala-1.15/>

Wang, J., & Gu, Q. (2024). Quantitative Assessment of Translation Quality in Education, Certification, and Industry: An Overview. *Open Journal of Modern Linguistics*, 14(2), 209-223. <https://doi.org/10.4236/ojml.2024.142012>

Wang, J., & Yu, X. (2025). Research on Automatic Evaluation Methods for English Translation Quality Integrated With Deep Learning. *IEEE Access*, 13, 212959-212972. <https://doi.org/10.1109/ACCESS.2025.3639245>

Wang, Y., Ma, X., Zhang, G., Ni, Y., Chandra, A., Guo, S., Ren, W., Arulraj, A., He, X., Jiang, Z., Li, T., Ku, M., Wang, K., Zhuang, A., Fan, R., Yue, X., & Chen, W. (2024). *MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark* (arXiv:2406.01574). arXiv. <https://doi.org/10.48550/arXiv.2406.01574>

Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). *Do Llamas Work in English? On the Latent Language of Multilingual Transformers* (arXiv:2402.10588). arXiv. <https://doi.org/10.48550/arXiv.2402.10588>

Wu, N. (2025). Beyond Prompt Engineering: The Evolution of Reasoning in Advanced Large Language Models. *International Journal on Science and Technology*, 16(1), 3719. <https://doi.org/10.71097/IJSAT.v16.i1.3719>

Xu, L., Xie, H., Qin, S.-Z. J., Tao, X., & Wang, F. L. (2023). *Parameter-Efficient Fine-Tuning Methods for Pretrained Language Models: A Critical Review and Assessment* (arXiv:2312.12148). arXiv. <https://doi.org/10.48550/arXiv.2312.12148>

Xu, Y., Hu, L., Zhao, J., Qiu, Z., XU, K., Ye, Y., & Gu, H. (2025). A Survey on Multilingual Large Language Models: Corpora, Alignment, and Bias. *Frontiers of Computer Science*, 19(11), 1911362. <https://doi.org/10.1007/s11704-024-40579-4>

Yan, J., Yan, P., Chen, Y., Li, J., Zhu, X., & Zhang, Y. (2024). *GPT-4 vs. Human Translators: A Comprehensive Evaluation of Translation Quality Across Languages, Domains, and Expertise Levels* (arXiv:2407.03658). arXiv. <https://doi.org/10.48550/arXiv.2407.03658>

Yang, W., & Yang, M. (2026). Evaluating literary translation by large language models: A multidimensional quality assessment of Shen Congwen's Border Town. *Humanities and Social Sciences Communications*, 13(1), 628. <https://doi.org/10.1057/s41599-026-06868-y>

Yenduri, G., Murugan, R., Govardanan, C., Supriya, Y., Srivastava, G., Reddy, P., Raj, D., Jhaveri, R., B, P., Wang, W., Vasilakos, A., & Gadekallu, T. (2024). GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access*, PP, 1-1. <https://doi.org/10.1109/ACCESS.2024.3389497>

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). A Survey on Multimodal Large Language Models. *National Science Review*, 11(12), nwae403. <https://doi.org/10.1093/nsr/nwae403>

Yong, Z. X., Schoelkopf, H., Muennighoff, N., Aji, A. F., Adelani, D. I., Almubarak, K., Bari, M. S., Sutawika, L., Kasai, J., Baruwa, A., Winata, G., Biderman, S., Raff, E., Radev, D., & Nikoulina, V. (2023). BLOOM+1: Adding Language Support to BLOOM for Zero-Shot Prompting. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Red.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 11682-11703). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.acl-long.653>

Yu, J., Bekerian, D., & Osback, C. (2024). Navigating the Digital Landscape: Challenges and Barriers to Effective Information Use on the Internet. *Encyclopedia*, 4, 1665-1680. <https://doi.org/10.3390/encyclopedia4040109>

Zhang, B., Cheng, Y., Shakeri, S., Wang, X., Ma, M., & Firat, O. (2025). *Encoder-Decoder or Decoder-Only? Revisiting Encoder-Decoder Large Language Model* (arXiv:2510.26622). arXiv. <https://doi.org/10.48550/arXiv.2510.26622>

Zhang, D., Song, J., Bi, Z., Song, X., Yuan, Y., Wang, T., Yeong, J., & Hao, J. (2025). *Mixture of Experts in Large Language Models* (arXiv:2507.11181). arXiv. <https://doi.org/10.48550/arXiv.2507.11181>

Zhang, M., Press, O., Merrill, W., Liu, A., & Smith, N. A. (2023). *How Language Model Hallucinations Can Snowball* (arXiv:2305.13534). arXiv. <https://doi.org/10.48550/arXiv.2305.13534>

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2025). Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *Computational Linguistics*, 51(4), 1373-1418. <https://doi.org/10.1162/COLI.a.16>

Zhong, T., Yang, Z., Liu, Z., Zhang, R., You, W., Liu, Y., Sun, H., Pan, Y., Li, Y., Zhou, Y., Jiang, H., Chen, J., & Liu, T. (2026). *Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research* (arXiv:2412.04497). arXiv. <https://doi.org/10.48550/arXiv.2412.04497>

Bijlagen

Bijlage A: Aangepaste MQM-hiërarchie met uitleg

MQM Categories	
<u>Accuracy</u>	Shifts occurring when the target content does not accurately correspond to the propositional content of the source text because of distortion, omission, or addition to the message.
Mistranslation	Shift occurring when the target content does not accurately represent the source content.
<i>Ambiguous target content</i>	Target content that introduces an ambiguity that is not present in the source content.
<i>False friend</i>	Incorrect use in target content of a word that is superficially similar to a source word.
<i>Overly literal</i>	Word-for-word equivalent in target content instead of idiomatic translation.
<i>MT hallucination</i>	Machine translation that is completely decoupled from the sense of the input sentence.
Overtranslation	Shift occurring in the target content that is inappropriately more specific than the source content.
Undertranslation	Shift occurring in the target content that is inappropriately less specific than the source content.
Addition	Shift occurring in the target content that includes content not present in the source.
Omission	Shift where content present in the source is missing in the target.
Unjustified euphemism	Target content that is potentially offensive in some way in the source language, but that has been inappropriately "watered down" in the translation.
<i>Obscenity</i>	A "watered down" representation of obscene content from the source text.
<i>Non-inclusivity</i>	An inaccurate or altered representation of source text content related to diversity, equity, or inclusion.
Untranslated	Shift occurring when a text segment that was intended for translation is left untranslated in the target content.
Culture-specific reference	An inaccurate or altered representation of culture-specific content within roughly the same semantical field as the source text.
<i>Religious content</i>	An inaccurate or altered representation of religious content from the source text.
Anachronism	Shift arising from the mention of a person, thing, or idea that exists out of its time in history, especially one that happened or existed later than the period being shown, discussed, etc.

MQM Categories <i>(vervolg)</i>	
<u>Linguistic conventions</u>	Shifts related to the linguistic well-formedness of the text, including problems with grammaticality, idiomaticity, and mechanical correctness.
Grammar	Shift that occurs when a text string violates the grammatical rules of the target language or unnecessarily changes the grammatical structure present in the source text.
Spelling	Shift occurring when a word is misspelled.
Duplication	Content repeated unintentionally.
Unintelligible	Text garbled or incomprehensible.
<u>Style</u>	Shifts occurring in a text that are grammatically acceptable but inappropriate because they deviate from style guides or exhibit inappropriate language style.
Awkward style	Style involving excessive wordiness or overly embedded clauses.
Unidiomatic style	Style that is grammatical but unnatural.
<u>Locale conventions</u>	Shifts occurring when the translation product violates locale-specific content or formatting requirements.
<u>Bracketed text</u>	Content that is bracketed.
Unnecessary for translation	Bracketed content that is unnecessary for a complete text (e.g. Synonyms, additional information, etc.).
Necessary for translation	Bracketed content that is necessary for a complete text.

Bijlage B: Handleiding voor de 'MQM annotation app'

1. Initiële configuratie en tekstinvoer

1. Tekst invoeren: Bij het openen van de applicatie verschijnt een invoerveld. Plak hier het te evalueren tekstfragment.
2. Bevestigen: Klik op de knop "Confirm Text & Start Annotating" om de annotatiemodus te activeren. *Let op*: Na bevestiging kan de basistekst niet meer worden gewijzigd zonder de sessie te herstarten (door de webpagina te vernieuwen).

2. Beheer van de taxonomie

De applicatie is standaard voorzien van de uitgebreide, taxonomie uit MQM Full. Deze kan volledig worden aangepast aan specifieke projectvereisten.

- Categorieën toevoegen: Gebruik de knop "[+] Add Base Category" onderaan het paneel voor een nieuwe hoofdcategorie. Om een subcategorie aan te maken, klikt u op "[+] Add child" onder een bestaande categorie. Nieuwe subcategorieën nemen automatisch de kleurcode van hun moedercategorie over.
- Titels wijzigen: Klik op het potloodicoon [✎] naast een categorienaam om deze te hernoemen.
- Definities wijzigen: Klik op het vraagtekenicoon [?] om de uitleg van de categorie te zien. In de pop-up kunt u op "Edit" klikken om deze beschrijving aan te passen.
- Categorieën verwijderen: Klik op het prullenbakicoon [🗑]. *Waarschuwing*: Het verwijderen van een categorie verwijdert tevens alle onderliggende subcategorieën én alle reeds gemaakte tekstannotaties die aan deze categorieën gekoppeld zijn.
- Taxonomie importeren/exporteren: Bovenin het categorieënpaneel kunt u via "Export Categories" uw op maat gemaakte taxonomie opslaan als een .json-bestand. Met "Import Categories" kunt u deze in een latere sessie hergebruiken.

3. Tekst annoteren

Het daadwerkelijke annotatieproces vindt plaats in de "Document Viewer".

1. Tekst selecteren: Selecteer met uw cursor het tekstsegment dat een verschuiving bevat.
2. Categorie toewijzen: Klik in het rechter- of linkerpaneel (afhankelijk van schermgrootte) op de gewenste categorie.
3. Ernstgraad (Severity) bepalen: Er verschijnt direct een pop-upvenster waarin u de ernst van de verschuiving moet specificeren:
 - Neutral (0 strafpunten)
 - Minor (1 strafpunt)
 - Medium (3 strafpunten)
 - Major (9 strafpunten)

Zodra u een keuze maakt, wordt de annotatie opgeslagen en gevisualiseerd.

4. Visualisatie en bewerking van annotaties

- Gelaagde structuur: Annotaties worden weergegeven als gekleurde balken onder de betreffende tekstsegmenten. Indien een tekstsegment meerdere verschuivingen bevat, worden deze overlappend gestapeld.
- Annotatie verwijderen: Indien u een foutieve annotatie wilt ongedaan maken, klikt u op de specifieke gekleurde laag onder de tekst. Er verschijnt een dialoogvenster (met de optie "[-] Remove Annotation") om deze actie te voltooien.

5. Kwantificering en Rapportage

Bovenaan de interface en in het zijpaneel worden de resultaten continu bijgehouden.

- Live Scorecard: Dit paneel toont real-time de cumulatieve punten en biedt een categorische uitsplitsing van het aantal geregistreerde verschuivingen.
- Sessie Opslaan / Laden: Met de knop "Save Session (.json)" in de bovenste navigatiebalk exporteert u de volledige werkomgeving (inclusief tekst, taxonomie en alle annotaties). U kunt dit bestand later inladen via de "Upload File" functionaliteit op het startscherm om uw werk te hervatten.
- Rapportage genereren: Klik op "Export Report (.txt)" om een overzichtelijk tekstdocument te genereren. Dit rapport bevat de kwantitatieve eindscores, een overzicht per categorie, een chronologische index van alle specifieke tekstverschuivingen, en de volledige originele brontekst ter referentie.

6. Resultaten Vergelijken

De applicatie beschikt over een geïntegreerde analysemodule genaamd 'Compare Results'. Deze module is specifiek ontworpen om de kwantitatieve resultaten van meerdere geëvalueerde teksten gelijktijdig te bestuderen en op te tellen. U opent dit onderdeel door rechtsonder in het scherm op het map-icoontje te klikken.

- Bestanden inladen: In het venster dat opent, kunt u via de selectieknop meerdere opgeslagen sessiebestanden (.json) vanaf uw computer tegelijkertijd uploaden.
- Vergelijkende weergave: De data uit de ingeladen bestanden wordt overzichtelijk naast elkaar in een tabel gepresenteerd. Per bestand ziet u de respectievelijke score en een uitsplitsing van de toegewezen ernstgraden.
- Totaalberekening: Aan de uiterste rechterzijde van de tabel genereert de applicatie automatisch een kolom waarin alle geregistreerde scores zijn opgeteld.
- Gedetailleerde navigatie: De tabel respecteert de hiërarchische structuur van uw taxonomie. Rechtsonder in het vak van een hoofdcategorie bevindt zich een klein pijltje. Door hierop te klikken vouwt u de specifieke subcategorieën open voor een gedetailleerder inzicht. Bovenaan in de kolom, bij 'Category', bevindt zich een pijltje waarmee u met één klik de gehele tabel volledig kunt uitklappen of inklappen.

Bijlage C: HTML-code voor 'MQM annotation app'

Door tweemaal op de onderstaande tekst te klikken, opent zich automatisch een document met de volledig HTML-code.

```
<!DOCTYPE html>
<html lang="nl">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>MQM Annotation App</title>

  <!-- Tailwind CSS -->
  <script src="https://cdn.tailwindcss.com"></script>

  <!-- React & ReactDOM -->
  <script crossorigin src="https://unpkg.com/react@18/umd/react.development.js"></script>
  <script crossorigin src="https://unpkg.com/react-dom@18/umd/react-dom.development.js"></script>

  <!-- Babel for transpiling JSX in the browser -->
  <script src="https://unpkg.com/@babel/standalone@7.21.4/babel.min.js"></script>

  <style>
    body { font-family: ui-sans-serif, system-ui, -apple-system, BlinkMacSystemFont, "Segoe UI",
    Roboto, "Helvetica Neue", Arial, sans-serif; }
  </style>
</head>
<body class="bg-slate-50 text-slate-900">
  <div id="root"></div>
```

Bijlage D: Statistieken van de drie bestudeerde Large Language Models

	TOTAAL (Gemini)	TOTAAL (Copilot)	TOTAAL (Kimi)	EINDTOTAAL
	29 pt. 20min, 3 med	124 pt. 25min, 18 med, 5 maj	106 pt. 22min, 16med, 4 maj	259 pt. 67min, 35med, 10maj
ACCURACY	22 pt. 16min, 2med	97 pt. 3neu, 19min, 14med, 4maj	74 pt. 1neu, 14min, 11med, 3maj	193 pt. 3neu, 49min, 25med, 8maj
Mistranslation	6 pt. 3min, 1med	36 pt. 6min, 7med, 1maj	52 pt. 7min, 6med, 3maj	94 pt. 16min, 14med, 4maj
<i>Ambiguous target content</i>	1 pt. 1min	1 pt. 1min	3 pt. 1med	5 pt. 2min, 1med
<i>False friend</i>	-	3 pt. 1med	-	3 pt. 1med
<i>Overly literal</i>	1 pt. 1min	18 pt. 3min, 5med	25 pt. 4min, 4med, 1maj	44 pt. 8min, 9med, 1maj
<i>MT hallucination</i>	-	-	-	-
Overtranslation	3 pt. 3min	-	2 pt. 2min	5 pt. 5min
Undertranslation	1 pt. 1min	11 pt. 2min, 3med	7 pt. 1min, 2med	19 pt. 4min, 5med
Addition	7 pt. 7min	9 pt. 3min, 2med	2 pt. 2min	18 pt. 12min, 2 med
Omission	1 pt. 1min	13 pt. 1min, 1med, 1maj	-	14 pt. 2min, 1med, 1maj
Unjustified euphemism	-	10 pt. 1min, 1maj	-	10 pt. 1min, 1maj
<i>Obscenity</i>	-	9 pt. 1maj	-	9 pt. 1maj
<i>Non-inclusivity</i>	-	1 pt. 1min	-	1 pt. 1min
Untranslated	-	9 pt. 1maj	6 pt. 1neu, 2med	15 pt. 1neu, 2med, 1maj
Culture-specific reference	3 pt. 1med	3 pt. 1med	4 pt. 1min, 1med	10 pt. 1min, 3med
<i>Religious content</i>	-	-	-	-
Anachronism	-	-	-	-

LINGUISTIC CONVENTIONS	3 pt. 3min	16 pt. 4min, 4med	14 pt. 2min, 4med	33 pt. 9min, 8med
Grammar	3 pt. 3min	13 pt. 4min, 3med	14 pt. 2min, 4med	30 pt. 9min, 7med
Spelling	-	-	-	-
Duplication	-	3 pt. 1med	-	3 pt. 1med
Unintelligible	-	-	-	-
STYLE	1 pt. 1min	11 pt. 2min, 1maj	18 pt. 6min, 1med, 1maj	30 pt. 9min, 1med, 2maj
Awkward style	-	-	-	-
Unidiomatic style	1 pt. 1min	-	6 pt. 3min, 1med	7 pt. 4min, 1med
LOCALE CONVENTIONS	-	-	-	-
BRACKETED TEXT	3 pt. 2neu, 1med	-	-	3 pt. 2neu, 1med
Unnecessary for translation	3 pt. 1neu, 1med	-	-	3 pt. 1neu, 1med
Necessary for translation	0 pt. 1neu	-	-	0 pt. 1neu

Bijlage E: Gemiddeld aantal punten per vers

Het aantal verzen van elk epigram is vermenigvuldigd met factor 3 in de tabel, aangezien het aantal punten van de drie LLM-vertalingen bij elkaar zijn opgeteld. Op die manier kan correct het gemiddelde worden berekend.

Epigram	Aantal verzen	Aantal punten	Gemiddeld aantal Punten per vers
<u>1.1</u>	18	22	1,22
<u>1.4</u>	24	54	2,25
<u>1.16</u>	6	2	0,33
<u>1.19</u>	12	5	0,42
<u>1.23</u>	12	1	0,08
<u>1.24</u>	12	14	1,17
<u>1.28</u>	6	6	1,00
<u>1.32</u>	6	5	0,83
<u>1.33</u>	12	6	0,50
<u>1.34</u>	30	65	2,17
<u>2.3</u>	6	3	0,50
<u>2.7</u>	24	10	0,42
<u>2.8</u>	24	27	1,13
<u>2.11</u>	30	34	1,13
<u>2.13</u>	6	5	0,83
<u>Totaal</u>	228	259	1,15

- Epigrammen met 2 (6) verzen: gemiddeld 0,70 punten per vers
- Epigrammen met 4 (12) verzen: gemiddeld 0,54 punten per vers
- Epigrammen met 6 (18) verzen: gemiddeld 1,22 punten per vers
- Epigrammen met 8 (24) verzen: gemiddeld 1,26 punten per vers
- Epigrammen met 10 (30) verzen: gemiddeld 1,65 punten per vers

Bijlage F: Gemiddeld aantal verschuivingen per vers

Het aantal verzen van elk epigram is vermenigvuldigd met factor 3 in de tabel, aangezien het aantal punten van de drie LLM-vertalingen bij elkaar zijn opgeteld. Op die manier kan correct het gemiddelde worden berekend.

Epigram	Aantal verzen	Shifts (totaal)	Gemiddeld aantal Shifts per vers
<u>1.1</u>	18	12	0,67
<u>1.4</u>	24	18	0,75
<u>1.16</u>	6	2	0,33
<u>1.19</u>	12	5	0,42
<u>1.23</u>	12	1	0,08
<u>1.24</u>	12	8	0,67
<u>1.28</u>	6	2	0,33
<u>1.32</u>	6	3	0,50
<u>1.33</u>	12	6	0,50
<u>1.34</u>	30	17	0,57
<u>2.3</u>	6	1	0,17
<u>2.7</u>	24	6	0,25
<u>2.8</u>	24	15	0,63
<u>2.11</u>	30	14	0,47
<u>2.13</u>	6	3	0,50
<u>Totaal</u>	228	113	0,50

- Epigrammen met 2 (6) verzen: gemiddeld 0,37 punten per vers
- Epigrammen met 4 (12) verzen: gemiddeld 0,42 punten per vers
- Epigrammen met 6 (18) verzen: gemiddeld 0,67 punten per vers
- Epigrammen met 8 (24) verzen: gemiddeld 0,54 punten per vers
- Epigrammen met 10 (30) verzen: gemiddeld 0,51 punten per vers

Bijlage G: Loeb-editie (1993) van de bestudeerde epigrammen

Epigram 1.1

*Hic est quem legis ille, quem requiris,
 toto notus in orbe Martialis
 argutis epigrammaton libellis:
 cui, lector studiose, quod dedisti
 viventi decus atque sentienti,
 rari post cineres habent poetae.*

Epigram 1.4

*Contigeris nostros, Caesar, si forte libellos,
 terrarum dominum pone supercilium.
 consuevere iocos vestri quoque ferre triumphi,
 materiam dictis nec pudet esse ducem.
 qua Thymelen spectas derisoremque Latinum,
 illa fronte precor carmina nostra legas.
 innocuos censura potest permittere lusus:
 lasciva est nobis pagina, vita proba.*

Epigram 1.16

*Sunt bona, sunt quaedam mediocria, sunt mala plura
 quae legis hic: aliter non fit, Avite, liber.*

Epigram 1.19

*Si memini, fuerant tibi quattuor, Aelia, dentes:
 expulit una duos tussis et una duos.
 iam segura potes totis tussire diebus:
 nil istic quod agat tertia tussis habet.*

Epigram 1.23

*Invitas nullum nisi cum quo, Cotta, lavis
 et dant convivam balnea sola tibi.
 mirabar quare numquam me, Cotta, vocasses:
 iam scio me nudum displicuisse tibi.*

Epigram 1.24

*Aspicias incompertis illum, Deciane, capillis,
 cuius et ipse times triste supercilium,
 qui loquitur Curios adsertoresque Camillos?
 nolito fronti credere: nupsit heri.*

Epigram 1.28

*Hesterno fetere mero qui credit Acerram,
 fallitur: in lucem semper Acerra bibit.*

Epigram 1.32

*Non amo te, Sabidi, nec possum dicere quare:
 hoc tantum possum dicere, non amo te.*

Epigram 1.33

*Amissum non flet cum sola est Gellia patrem;
 si quis adest, iussae prosiliunt lacrimae.
 non luget quisquis laudari, Gellia, quaerit;
 ille dolet vere qui sine teste dolet.*

Epigram 1.34

*Incustoditis et apertis, Lesbia, semper
 liminibus peccas nec tua furta tegis,
 et plus spectator quam te delectat adulter
 nec sunt grata tibi gaudia si qua latent.
 at meretrix abigit testem veloque seraque
 raraque Summemmi fornice rima patet.
 A Chione saltem vel ab Iade disce pudorem:
 abscondunt spurcas et monumenta lupas.
 numquid dura tibi nimium censura videtur?
 deprende veto te, Lesbia, non futui.*

Epigram 2.3

*Sexte, nihil debes, nil debes, Sexte, fatemur.
 debet enim, si quis solvere, Sexte, potest.*

Epigram 2.7

*Declamas belle, causas agis, Attale, belle,
 historias bellas, carmina bella facis,
 componis belle mimos, epigrammata belle,
 bellus grammaticus, bellus es astrologus,
 et belle cantas et saltas, Attale, belle,
 bellus es arte lyrae, bellus es arte pilae.
 nil bene cum facias, facias tamen omnia belle,
 vis dicam quid sis? magnus es ardalio.*

Epigram 2.8

*Si qua videbuntur chartis tibi, lector, in istis
 sive obscura nimis sive Latina parum,
 non meus est error: nocuit librarius illis
 dum properat versus annumerare tibi.
 quod si non illum sed me peccasse putabis,
 tunc ego te credam cordis habere nihil.
 'ista tamen mala sunt.' quasi nos manifesta negemus!
 haec mala sunt, sed tu non meliora facis.*

Epigram 2.11

*Quod fronte Selium nubila vides, Rufe,
 quod ambulator porticum terit seram,
 lugubre quiddam quod tacet piger vultus,
 quod paene terram nasus indecens tangit,
 quod dextra pectus pulsant et comam vellit:
 non ille amici fata luget aut fratris,
 uterque natus vivit et precor vivat,
 salva est et uxor sarcinaeque servique,
 nihil colonus vilicusque decoxit.
 maeroris igitur causa quae? domi cenat.*

Epigram 2.13

*Et iudex petit et petit patronus.
 solvas censeo, Sexte, creditori.*

Bijlage H: Vertalingen door Piet Schrijvers (2019)

Epigram 1.1

Hier is hij die je leest en graag wilt kopen:
Martialis, bekend over de hele wereld
om zijn puntige bundels epigrammen.
Beste lezer, de roem die je hem verleende
bij zijn leven en bewustzijn, krijgen
dichters zelden als ze gecremeerd zijn.

Epigram 1.4

Keizer, wanneer u toevallig onze boekjes ter hand neemt,
ontspan dan uw frons, die over de aarde heerst.
Ook uw triomfen zijn gewend om spot te verdragen,
een generaal schaamt zich niet stof voor grappen te zijn.
Met het gezicht waarmee u kijkt naar ballet of een mime,
moet u mijn epigrammen lezen, zo bid ik u.
Een censor kan aan onschuldige scherts zijn toestemming geven.
Mijn pagina is gewaagd, mijn levenswijze fatsoenlijk.

Epigram 1.16

Sommige stukken zijn goed, andere middelmatig, de meeste
slecht van wat je hier leest. Zo ontstaat nu eenmaal een boek.

Epigram 1.19

Als ik me goed herinner, bezat jij, Aelia, vier tanden;
één hoestbui nam er twee weg, een tweede hoestbui de rest.
Nu kun je onbezorgd de hele dag door blijven hoesten:
een derde hoestbui die volgt, heeft immers niets meer te doen.

Epigram 1.23

Cotta, jij nodigt slechts mensen uit met wie je kunt baden.
Louter het badhuis levert aan jou een gast.
Ik vroeg me al af waarom jij, Cotta, mij nooit inviteerde?
Nu weet ik het: naakt viel ik bij jou niet in de smaak.

Epigram 1.24

Zie je die man, Decianus, met die verwaarloosde haren,
door wiens grimmige frons jij wordt geïmponeerd,
die zijn mond vol heeft van Curius en andere vrijheidsstrijders?
Vergis je niet in die aanblik. Gisteren was hij de bruid!

Epigram 1.28

Wie gelooft dat Acerra stinkt naar de wijn van gisteren,
vergist zich: Acerra drinkt steeds tot het ochtenduur.

Epigram 1.32

Ik mag je niet, Sabidius, maar waarom? Dat kan ik niet zeggen.
Ik kan alleen maar zeggen, Sabidius: ik mag je niet.

Epigram 1.33

Gellia beweent haar gestorven vader niet als ze alleen is,
maar als iemand erbij is, springen tranen tevoorschijn.
Gellia, niet werkelijk rouwt al wie geprezen wil worden.
Slechts diegene treurt echt die zonder getuige treurt.

Epigram 1.34

Altijd is jouw deur onbewaakt en staat open, Lesbia,
als je vreemdgaat en niet je overspel bedekt.
Je hebt meer plezier in een toeschouwer dan in een minnaar,
genoegens bevallen jou niet als ze verborgen zijn.
Maar een prostituee verwijdt met gordijn en grendel
iedere getuige, geen spleet staat open in een bordeel.
Leer toch tenminste van Chia of las gevoelens van schaamte:
grafmonumenten verbergen de allersmerigste hoer.
Lesbia, getuigt volgens jou mijn kritiek van te grote strengheid?
Je mag worden geneukt, als je maar niet wordt betraapt.

Epigram 2.3

Sextus, ik moet het erkennen, jij bent mij werkelijk niets schuldig.
Om iets schuldig te zijn, moet je, Sextus, iets kunnen betalen.

Epigram 2.7

Je declameert heel aardig, Atticus, je pleidooien zijn aardig,
je schrijft aardige stukken, en aardig is menig gedicht.
Je maakt aardige mimen, aardige epigrammen,
je bent een aardige literator en astronoom.
Je hebt een aardige zangstem, Atticus, je danst ook aardig.
Je bespeelt aardig de lier en aardig speel je met de bal.
Terwijl je niets goed doet, doe je toch alles aardig.
Zal ik zeggen wat je bent? Een bedrijvige nietsnut ben jij!

Epigram 2.8

Lezer, wanneer je vindt dat in deze epigrammen
soms iets te duister is, soms iets niet goed Latijn,
ligt de fout niet bij mij. De kopiist heeft ze bedorven,
toen hij overhaastig versregels afschreef voor jou.
Als je denkt dat niet hij, maar dat ik in de fout was,
zal ik denken dat je er geen verstand van hebt.
'Toch zijn ze slecht', alsof ik ontken wat evident is.
Deze zijn slecht, maar jij schrijft toch geen betere!

Epigram 2.11

Dat Selius met bewolkt gezicht passeert, Rufus,
en nog zo laat de zuilencolonnade platloopt,
dat zijn gelaat zo lijdzaam is en zwijgzaam,
zijn neus zo onbehoorlijk bijna de grond raakt,
dat hij op zijn borst slaat en zijn haren uittrekt:
hij treurt niet om een vriend of broer die dood is,
zijn beide zonen leven en, zo bid ik, blijven leven.
Zijn vrouw, zijn meubels en zijn slaven zijn in orde.
Geen opzichter, geen pachter die hem wil bestelen.
Wat is de oorzaak van zijn droefheid? Hij eet thuis.

Epigram 2.13

Een rechter vraagt om geld, een advocaat vraagt geld.
Sextus, je kunt maar beter je schuldeiser betalen.

Bijlage I: Vertalingen door Frans Van Dooren (1996)

Epigram 1.1

Hier is dan die bekende Martialis,
die men graag leest omdat hij speels van taal is
en scherpe epigrammen heeft geschreven.
De grote roem die hij al bij zijn leven
van u, o trouwe lezer, mag verwerven,
krijgen veel dichters niet eens na hun sterven.

Epigram 1.4

O keizer, als u een gedicht
van mij onder ogen mocht komen,
ontspan dan uw strenge gezicht,
symbool van de almacht van Rome.
Kwinkslagen en grappen en steken
zijn bij uw triomfen gewoon:
geen veldheer of vorst zal verbleken
bij spot op zijn eigen persoon.
Dus lees mijn gedichtenboek zó
zoals u naar mimen gaat kijken
of wil het geheel met een show
vol scherpe satire vergelijken.
onschuldige scherts kunt u rustig,
o keizer, als censor vergeven:
mijn boeken zijn speels en wellustig,
maar eerbaar en kuis is mijn leven.

Epigram 1.16

Al wat je hier kunt lezen
is goed of slecht of gaat,
omdat alleen op deze
manier een boek ontstaat.

Epigram 1.19

Zeg Aelia, al roep je ach en wee,
naar je vier tanden kun je fluiten:
je hoestte er eerst de eerste twee
en toen de laatste twee naar buiten.
Maar weet je wat het leed verzoet
nu je die bijters hebt verloren?
Dat jij, als je wéér hoesten moet,

je nergens meer aan hoeft te storen.

Epigram 1.23

Jij vraagt slechts hen bij je te gast
die jou, o Cotta, in de thermen,
als jij je 't zweet van 't lichaam wast,
driftig belagen en omzwermen.
En was ik eerst nog wat verrast
dat jij mij nooit voor 't maal kwam vragen,
nu staat één ding voor mij wel vast:
dat ik je naakt niet kan behagen.

Epigram 1.24

Zeg Decianus, zie jij daar
die kerel met die woeste blikken
en met dat ongekamde haar,
van wie de mensen steeds weer schrikken
als hij zijn schuifdeur opendoet
en onbekookt begint te schreeuwen
over de kracht en deugd en moed
van helden uit voorbije eeuwen?
Welnu, al kraamt hij dat niet uit,
men nam hem gisteren tot bruid.

Epigram 1.28

Hoe kan Acerra nou
naar wijn van gisteren stinken
wanneer Acerra trouw
tot 's morgens door blijft drinken?

Epigram 1.32

Sabidius, ik moet jou niet!
Maar als ik het je uit moet leggen,
dan kan ik dit alleen maar zeggen:
Sabidius, ik moet jou niet!

Epigram 1.33

Als Gellia alleen is, treurt ze niet
om vader die het leven liet,
terwijl ze traan na traan vergiet
zo gauw er iemand binnenschiet.

Maar Gellia, als je zo graag ziet
dat met men je prijs, dan rouw je niet.
Want zij slechts hebben echt verdriet,
die schreien als geen mens het ziet.

Epigram 1.34

O Lesbia, jij hoereert altijd
met onbewaakte open deuren,
en steeds als jij je benen spreidt
kan iedereen dat zien gebeuren.
Een man die kijkt bevalt jou meer
dan een die langs je billen wrijft,
en je geniet pas echt wanneer
de ontucht niet verborgen blijft.
Maar hoeren sluiten tegen 't gluren
de deur met grendels en gordijnen,
en in geen krothuis bij de muren
zie je het zonlicht binnenschijnen.
't Wordt tijd dat jij je gaat generen
en doet zoals je concurrenten
van wie er zelfs prostitueren
in graftomben en monumenten.
En om je nu mijn streng beklag
eens en voorgoed te laten snappen
zeg ik je, Lesbia, neuken mág
maar laat geen mens je erbij betrappen!

Epigram 2.3

Je hebt geen schuld Sextus, geen cent.
Je hebt geen schuld, dat is bekend.
Want schuld, daar spreek je alleen maar van
bij iemand die betalen kan.

Epigram 2.7

Aticus, je declameert heel aardig
en voor de balie ben je best slagvaardig.
Aardig schrijf je over de historie
en ook de poëzie verschaft je glorie.
Aardig zijn de mimen die je dicht
en al je epigrammen zijn gericht.
Aardig ben je meester van de taal

en van de astrologie al helemaal.
Aardig, Atticus, kun je ook zingen
en aardig kun je op de dansvloer springen.
Aardig maak je op de lier muziek
en aardig is bij 't balspel je techniek.
En al die je in feite niet echt goed,
't is allemaal toch aardig wat je doet.
Wat ik van je vind? Wou je dat weten?
Ik vind je een aardig arrogant stuk vreten!

Epigram 2.8

Als mijn gedichten jou misschien
soms hier en daar wat duister zijn
en opgeschreven bovendien
in een wat minder goed Latijn,
dan is dat, lezer, niet mijn schuld,
want hij die ze overschrijft maakt haast
en zit er door jouw ongeduld
van tijd tot tijd wel eens naast.
Maar als hij desondanks toch vindt
dat ik die fouten heb gemaakt,
dan wordt het hoogste tijd, beste vrind, [*]
dat je uit je molleslaap ontwaakt.
'Toch zijn ze slecht, jouw verzen!' Och,
de waarheid moet zich maar ontvouwen:
natuurlijk zijn ze slecht, maar toch
nog altijd beter dan de jouwe.

[] 'vrind' is letterlijk overgenomen uit de vertaling van Frans Van Dooren (1996)*

Epigram 2.11

Waarom sluipt SeliuS nog zo laat
zwaar zuchtend door de zuilengangen?
Waarom laat hij, waar hij ook gaat,
zijn peinzend hoofd zo treurig hangen?
Waarom toch zou hij met zijn neus
de stenen van de grond haast raken?
Waarom zou hij triest en nerveus
in stilte jammerklachten slaken?
Is SeliuS soms in diepe rouw
vanwege een vriend of een bekende?

Is soms de dood van broer of vrouw
de oorzaak van zijn doffe ellende?
Is zijn bezit eraan gegaan
of zijn zijn slaven soms gestorven?
Is door de rentmeester het graan
op 't land verwaarloosd en bedorven?
Och nee, Rufus, maar wil je graag
de bron van zijn misère weten:
wel, dat is dit, dat hij vandaag
genoodzaakt is om thuis te eten.

Epigram 2.13

Als advocaat en rechter
hun centen komen halen,
dan is dat meestal slechter
dan zelf je schuld betalen.

Bijlage J: Annotaties van de vertalingen door Gemini

MQM BATCH ANNOTATION REPORT

=====

BESTAND: 1.1 MQM gemini .json

1. "Dit"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Medium (3 pts)

2. "beroemd"

Categorie : ACCURACY > Overtranslation

Ernstgraad: Minor (1 pts)

3. "een roem"

Categorie : ACCURACY > Addition

Ernstgraad: Minor (1 pts)

4. "na hun dood"

Categorie : ACCURACY > Undertranslation

Ernstgraad: Minor (1 pts)

5. "ten deel valt"

Categorie : LINGUISTIC CONVENTIONS > Grammar

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Dit is hij die je leest, naar wie je zoekt,
Martialis, in de hele wereld beroemd
om zijn geestige boekjes met epigrammen:
aan wie jij, toegewijde lezer, de roem hebt geschonken
terwijl hij nog leeft en het kan ervaren,
een roem die slechts weinig dichters na hun dood ten deel valt.

BESTAND: 1.4 MQM Gemini.json

1. "met diezelfde gezichtsuitdrukking"

Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

2. "best"

Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

3. "bladzijden zijn"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als u, Caesar, toevallig mijn boekjes in handen krijgt,
laat dan uw frons, die de wereld regeert, varen.
Ook uw triomftochten zijn gewend om grappen te verdragen,
en een leider schaamt zich er niet voor het onderwerp van spot te zijn.
Met de blik waarmee u naar Thymele en de komiek Latinus kijkt,
met diezelfde gezichtsuitdrukking, zo smee ik u, lees mijn gedichten.
De censor kan onschuldig vermaak best toestaan:
mijn bladzijden zijn losbandig, maar mijn leven is eerzaam.

BESTAND: 1.16 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

Sommige zijn goed, sommige zijn matig, de meeste die je hier leest zijn slecht: op een andere manier, Avitus, komt een boek niet tot stand.

BESTAND: 1.19 MQM Gemini.json

1. "hoestbui"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Minor (1 pts)

2. "en nóg een hoestbui twee"

Categorie : STYLE > Unidiomatic style
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als ik het me goed herinner, had je vier tanden, Aelia:
één hoestbui wierp er twee uit, en nóg een hoestbui twee.
Nu kun je zonder zorgen de hele dag door hoesten:
een derde hoestbui heeft daar niets meer uit te richten.

BESTAND: 1.23 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

Je nodigt niemand uit, Cotta, behalve wie met je in bad gaat,
en alleen het badhuis levert jou een tafelgenoot op.
Ik vroeg me al af waarom je me nooit had uitgenodigd, Cotta:
nu weet ik dat ik je naakt ben tegengevallen.

BESTAND: 1.24 MQM Gemini.json

1. "blik"

Categorie : ACCURACY > Overtranslation
Ernstgraad: Minor (1 pts)

2. "terugdeinst"

Categorie : ACCURACY > Overtranslation
Ernstgraad: Minor (1 pts)

3. "(oer-Romeinse helden als)"

Categorie : BRACKETED TEXT > Unnecessary for translation
Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Zie je die man daar met zijn warrige haar, Decianus,
voor wiens norske blik zelfs jij terugdeinst,
die de mond vol heeft van (oer-Romeinse helden als) de Curii en de bevrijders, de Camilli?
Laat je niet misleiden door die façade: hij is gisteren als bruid getrouwd.

BESTAND: 1.28 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

"Wie denkt dat Acerra naar de wijn van gisteren stinkt, vergist zich:
Acerra drinkt altijd door tot de ochtend."

BESTAND: 1.32 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

"Ik houd niet van jou, Sabidius, en ik kan niet zeggen waarom:
dit is het enige dat ik kan zeggen, ik houd niet van jou."

BESTAND: 1.33 MQM Gemini.json

1. "is er"
Categorie : ACCURACY > Omission
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Gellia huilt niet om haar overleden vader wanneer ze alleen is;
is er iemand aanwezig, dan springen haar tranen op commando tevoorschijn.
Wie geprezen wil worden, Gellia, rouwt niet;
wie zonder getuige treurt, die treurt pas echt.

BESTAND: 1.34 MQM Gemini.json

1. "slippertjes"

Categorie : ACCURACY > Mistranslation > Ambiguous target content

Ernstgraad: Minor (1 pts)

2. "zelf"

Categorie : ACCURACY > Addition

Ernstgraad: Minor (1 pts)

3. "zijn niets voor jou"

Categorie : ACCURACY

Ernstgraad: Minor (1 pts)

4. "oordeel"

Categorie : ACCURACY > Culture-specific reference

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Lesbia, je zondigt altijd met onbewaakte en open deuren,
en je verbergt je slippertjes niet.

Een toeschouwer geeft je meer genot dan de minnaar zelf,
en pleziertjes zijn niets voor jou als ze verborgen blijven.

Maar een prostituee weert getuigen met gordijn en grendel,
en in het bordeel van Summemmus kiert er zelden een spleet.

Leer tenminste wat schaamte van Chione of van Ias:

zelfs grafmonumenten houden smerige hoeren verborgen.*

Lijkt mijn oordeel je te streng?

Ik verbied je om betrappt te worden, Lesbia, niet om geneukt te worden.

BESTAND: 2.3 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

"Sextus, je bent niets schuldig, je hebt geen enkele schuld, Sextus, dat geven we toe.
Iemand heeft immers pas een schuld, Sextus, als hij in staat is te betalen."

BESTAND: 2.7 MQM Gemini.json

1. "écht"

Categorie : ACCURACY > Addition

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Je declameert aardig, je bepleit rechtszaken, Attalus, aardig,
aardige verhalen en aardige gedichten schrijf je,
je componeert aardig kluchten en epigrammen aardig,
een aardige taalkundige en een aardige astroloog ben je,
en je zingt aardig en je danst, Attalus, aardig,
je bent aardig in het lierenspel, aardig in het balspel.
Omdat je niets écht goed doet, maar toch alles aardig,
wil je dat ik zeg wat je bent? Je bent een grote dilettant.

BESTAND: 2.8 MQM Gemini.json

1. "(gebrekkig Latijn)"

Categorie : BRACKETED TEXT > Unnecessary for translation

Ernstgraad: Neutral (0 pts)

2. "maakt"

Categorie : ACCURACY > Mistranslation > Overly literal

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als er in deze bladzijden, lezer, dingen aan jou verschijnen
die ofwel te duister, ofwel te weinig Latijn (gebrekkig Latijn) zijn,
dan is dat niet mijn fout: de kopiist heeft ze verknoeid
terwijl hij zich haastte om de verzen voor jou af te tellen.
Maar als je zult denken dat niet hij, maar ik de fout heb gemaakt,
dan zal ik geloven dat jij geen greintje verstand hebt.
'Toch zijn ze slecht.' Alsof wij het overduidelijke zouden ontkennen!
Deze zijn slecht, maar jij maakt geen betere.

BESTAND: 2.11 MQM Gemini.json

1. "in een treurig zwijgen is gehuld"
Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)
2. "heus"
Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)
3. "(en ik bid dat ze in leven blijven)"
Categorie : BRACKETED TEXT > Necessary for translation
Ernstgraad: Neutral (0 pts)
4. "allemaal"
Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Dat jij Selius met een sombere blik ziet, Rufus,
dat hij wandelend tot laat de zuilengang verslijt,
dat zijn stuurse gezicht in een treurig zwijgen is gehuld,
dat zijn lelijke neus bijna de grond raakt,
dat zijn rechterhand op zijn borst slaat en haren uittrekt:
hij rouwt heus niet om het lot van een vriend of een broer.
Zijn beide zonen leven (en ik bid dat ze in leven blijven),
zijn vrouw, bezittingen en slaven zijn allemaal ongedeerd,
zijn pachter en bedrijfsleider hebben niets verbrast.
Wat is dan de oorzaak van zijn verdriet? Hij dineert thuis.

BESTAND: 2.13 MQM Gemini.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

Zowel de rechter als de advocaat eist geld.
Ik raad je aan, Sextus, om gewoon je schuldeiser te betalen.

=====

Bijlage K: Annotaties van de vertalingen door Copilot

MQM BATCH ANNOTATION REPORT

=====

BESTAND: 1.1 MQM Copilot VLOT.json

1. "Dit"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Medium (3 pts)

2. "wat je schonk"

Categorie : ACCURACY > Addition
Ernstgraad: Medium (3 pts)

3. "na hun dood"

Categorie : ACCURACY > Undertranslation
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Dit is hij die je leest, hij die je zoekt:
Martialis, over de hele wereld bekend
om zijn geestige bundeltjes epigrammen.
Aan hem heb jij, ijverige lezer,
wat je schonk—roem terwijl hij leefde en nog voelde—
iets gegeven wat maar weinig dichters na hun dood nog krijgen.

BESTAND: 1.4 MQM Copilot.json

1. "je"

Categorie : STYLE
Ernstgraad: Major (9 pts)

2. "vorstelijke"

Categorie : ACCURACY > Omission
Ernstgraad: Medium (3 pts)

3. "kunnen tegen"

Categorie : ACCURACY
Ernstgraad: Minor (1 pts)

4. "levert"

Categorie : ACCURACY > Mistranslation > Ambiguous target content
Ernstgraad: Minor (1 pts)

5. "zonder schaamte"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

6. "Lees"

Categorie : ACCURACY > Omission
Ernstgraad: Minor (1 pts)

7. "censuur"

Categorie : ACCURACY > Culture-specific reference
Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Wanneer je, Caesar, toevallig mijn boekjes oppakt,
zet dan even de vorstelijke frons af.
Zelfs jouw triomfen kunnen tegen een grap;
en de leider zelf levert zonder schaamte onderwerp voor spot.
Lees mijn gedichten met dezelfde glimlach
waarmee je Thymele en de grappenmaker Latinus bekijkt.
Want de censuur mag onschuldige spelerei toelaten:
mijn pagina is vrijpostig, maar mijn leven is deugdzaam.

BESTAND: 1.16 MQM Copilot VLOT.json

1. "maar"

Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

2. "Zo gaat dat nu eenmaal"

Categorie : ACCURACY > Undertranslation
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Wat je hier leest: sommige zijn goed, een paar middelmatig, maar de meeste slecht.
Zo gaat dat nu eenmaal, Avitus, met een boek.

BESTAND: 1.19 MQM Copilot.json

1. "zonder zorgen"
Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als ik het me goed herinner, Aelia, had jij vier tanden:
één hoest stootte er twee uit, en nog één hoest nog eens twee.
Nu kun je gerust de hele dag hoesten, zonder zorgen:
daar is niets meer dat een derde hoest kan uitrichten.

BESTAND: 1.23 MQM Copilot VLOT.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

Je nodigt niemand uit, Cotta, behalve wie met je gaat baden,
en alleen het badhuis bezorgt jou een tafelgenoot.
Ik vroeg me af waarom je mij nooit uitnodigde, Cotta:
nu weet ik het — ik heb je naakt niet bekoord.

BESTAND: 1.24 MQM Copilot VLOT.json

1. "sombere"
Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)
2. "deugdhelden"
Categorie : ACCURACY > Mistranslation
Ernstgraad: Minor (1 pts)

3. "voorhoofd"

Categorie : ACCURACY > Mistranslation > Overly literal

Ernstgraad: Minor (1 pts)

4. "getrouwd."

Categorie : ACCURACY > Undertranslation

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Je ziet hem daar, Decianus, met zijn onverzorgde haar—
die man van wie zelfs jij die sombere frons vreest,
die maar blijft praten over de Curii en de Camilli, die deugdhelden.
Geloof zijn voorhoofd niet: gisteren is hij getrouwd.

BESTAND: 1.28 MQM Copilot VLOT.json

1. "ruikt "

Categorie : ACCURACY > Undertranslation

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

“Wie meent dat Acerra ruikt naar de drank van gisteravond, zit mis: Acerra drinkt stevast
door tot het daglicht.”

BESTAND: 1.32 MQM Copilot VLOT.json

1. "wél"

Categorie : ACCURACY

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Ik hou niet van je, Sabidius, en ik kan niet zeggen waarom;
dit ene kan ik wél zeggen: ik hou niet van je.

BESTAND: 1.33 MQM Copilot VLOT.json

1. "opgedragen"

Categorie : STYLE

Ernstgraad: Minor (1 pts)

2. "hij"

Categorie : ACCURACY > Unjustified euphemism > Non-inclusivity

Ernstgraad: Minor (1 pts)

3. "waarlijk"

Categorie : STYLE

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Gellia huilt niet om haar gestorven vader wanneer ze alleen is;
maar zodra er iemand bij is, springen de opgedragen tranen tevoorschijn.
Wie rouwt om geprezen te worden, Gellia, rouwt niet echt;
pas hij treurt waarlijk, die treurt zonder getuige.

BESTAND: 1.34 MQM Copilot.json

1. "Sorry, dat kan ik niet doen"

Categorie : ACCURACY > Untranslated

Ernstgraad: Major (9 pts)

2. "geen vertaling of weergave "

Categorie : ACCURACY > Omission

Ernstgraad: Major (9 pts)

3. "seksueel expliciete inhoud "

Categorie : ACCURACY > Unjustified euphemism > Obscenity

Ernstgraad: Major (9 pts)

--- ORIGINELE TEKST ---

Sorry, dat kan ik niet doen. De tekst die je stuurde bevat expliciet seksuele taal (o.a. een grove term in de laatste regel), en ik mag geen vertaling of weergave van seksueel expliciete inhoud leveren. [...]

BESTAND: 2.3 MQM Copilot VLOT.json

1. "iemand"

Categorie : LINGUISTIC CONVENTIONS > Grammar

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Sextus, je bent niets schuldig—niets schuldig, Sextus, dat geven we toe.
Want schuldig is hij pas, als iemand het kan terugbetalen, Sextus

BESTAND: 2.7 MQM Copilot.json

1. "mooi;"

Categorie : LINGUISTIC CONVENTIONS > Duplication

Ernstgraad: Medium (3 pts)

2. "echt"

Categorie : ACCURACY > Addition

Ernstgraad: Minor (1 pts)

3. "druktemaker"

Categorie : ACCURACY

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Je declameert mooi, je bepleit zaken, Attalus, mooi;
je maakt mooie verhalen, mooie gedichten.
Je schrijft mooi toneelstukjes, mooie epigrammen;
je bent een knappe grammaticus, een knappe astroloog.
En mooi zing je en dans je, Attalus, mooi;
knap ben je in de kunst van de lier, knap in het spel met de bal.
Al doe je niets echt goed, toch doe je alles "mooi":
wil je dat ik zeg wat je bent? Een geweldige druktemaker ben je.

BESTAND: 2.8 MQM Copilot.json

1. "iets staat
dat"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

2. "weinig"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Minor (1 pts)

3. "nummeren"

Categorie : ACCURACY > Mistranslation > False friend
Ernstgraad: Medium (3 pts)

4. "gezondigd"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)

5. "hart "

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)

6. "het is"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

7. "het is"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

8. "maakt"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Minor (1 pts)

9. "het niet beter"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Als er op deze bladen, lezer, iets staat
dat je óf al te duister lijkt óf weinig Latijn,
dan is dat niet mijn fout: de kopiist heeft het bedorven
toen hij zich haastte de verzen voor jou te nummeren.
Maar als je denkt dat niet hĳ, maar ik heb gezondigd,

dan zal ik geloven dat jij geen hart hebt.
"Maar toch: het is slecht." Alsof wij het overduidelijke ontkennen!
Ja, het is slecht — maar jij maakt het niet beter.

BESTAND: 2.11 MQM Copilot.json

1. "Omdat"
Categorie : ACCURACY
Ernstgraad: Minor (1 pts)
2. "somber,"
Categorie : ACCURACY > Addition
Ernstgraad: Medium (3 pts)
3. "voorhoofd"
Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Minor (1 pts)
4. "omdat"
Categorie : ACCURACY
Ernstgraad: Minor (1 pts)
5. "late zuilengang"
Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)
6. "omdat"
Categorie : ACCURACY
Ernstgraad: Neutral (0 pts)
7. " zijn trage gezicht zwijgend iets rouwigs heeft"
Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Medium (3 pts)
8. "trage"
Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)
9. "omdat"
Categorie : ACCURACY
Ernstgraad: Neutral (0 pts)

10. "omdat"

Categorie : ACCURACY

Ernstgraad: Neutral (0 pts)

11. "ouders"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Major (9 pts)

--- ORIGINELE TEKST ---

Omdat je Selius met een somber, bewolkt voorhoofd ziet, Rufus,
omdat hij als wandelaar de late zuilengang afslijt,
omdat zijn trage gezicht zwijgend iets rouwigs heeft,
omdat zijn onbetamelijke neus bijna de grond raakt,
omdat hij met zijn rechterhand op zijn borst slaat en zijn haar uittrekt—
hij beweent niet de dood van een vriend of een broer:
beide ouders leven nog, en ik bid dat ze blijven leven;
ook zijn vrouw is ongedeerd, zijn have en zijn slaven eveneens;
geen pachter of rentmeester heeft iets verbrast.
Wat is dan de oorzaak van zijn verdriet? Hij eet thuis.

BESTAND: 2.13 MQM Copilot VLOT.json

1. "erom"

Categorie : ACCURACY > Undertranslation

Ernstgraad: Medium (3 pts)

2. "Volgens mij moet je betalen"

Categorie : ACCURACY

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

"Zowel de rechter als de advocaat vraagt erom.
Volgens mij moet je betalen, Sextus, aan de schuldeiser."

=====

Bijlage L: Annotaties van de vertalingen door Kimi

MQM BATCH ANNOTATION REPORT

=====

BESTAND: 1.1 MQM Kimi.json

1. "Dit"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Medium (3 pts)

2. "eer"

Categorie : ACCURACY > Culture-specific reference
Ernstgraad: Minor (1 pts)

3. "zeldzame"

Categorie : STYLE > Unidiomatic style
Ernstgraad: Medium (3 pts)

4. "as"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Dit is hij die je leest, die je zoekt,
over de hele wereld bekend, Martialis,
door zijn scherpe boekjes met epigrammen:
aan wie jij, ijverige lezer, wat je ook gaf
aan eer terwijl hij leefde en voelde,
dat hebben zeldzame dichters na hun as.

BESTAND: 1.4 MQM Kimi.json

1. "je"

Categorie : STYLE
Ernstgraad: Major (9 pts)

2. "heerser over de wereld"

Categorie : ACCURACY > Mistranslation > Ambiguous target content

Ernstgraad: Medium (3 pts)

3. "het is geen schande"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Minor (1 pts)

4. "Waar"

Categorie : LINGUISTIC CONVENTIONS > Grammar

Ernstgraad: Medium (3 pts)

5. "Latijnse"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Major (9 pts)

6. "daarop"

Categorie : LINGUISTIC CONVENTIONS > Grammar

Ernstgraad: Medium (3 pts)

7. "bid"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Minor (1 pts)

8. "censuur"

Categorie : ACCURACY > Culture-specific reference

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Caesar, als je misschien toevallig onze boekjes aanraakt,
leg dan je frons neer, heerser over de wereld.

Ook jouw triomfen hebben gewend grappen te verdragen,
en het is geen schande dat een leider stof voor spot biedt.

Waar je naar Thymele en de Latijnse spotvogel kijkt,
daarop, bid ik, lees mijn gedichten met diezelfde blik.

De censuur kan onschuldige spelletjes toestaan:
onze pagina is wellustig, ons leven is deugdzaam.

BESTAND: 1.16 MQM Kimi VLOT.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

"Er zijn goede gedichten, er zijn middelmatige, er zijn meer slechte die je hier leest: anders wordt het, Avitus, geen boek."

BESTAND: 1.19 MQM Kimi.json

1. "en één andere twee"

Categorie : STYLE > Unidiomatic style

Ernstgraad: Minor (1 pts)

2. "te zoeken"

Categorie : ACCURACY > Undertranslation

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als ik me goed herinner, Aelia, had jij vier tanden:
één hoestbui spuugde er twee uit, en één andere twee.
Nu kun je de hele dag onbezorgd doorhoesten:
daar heeft een derde hoestbui niets meer te zoeken.

BESTAND: 1.23 MQM Kimi.json

1. "de baden"

Categorie : ACCURACY > Mistranslation > Overly literal

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Je nodigt nooit iemand uit, Cotta, tenzij met wie je badt,
en alleen de baden geven jou een gast.
Ik vroeg me af waarom je me nooit uitnodigde, Cotta:
nu weet ik dat ik jou naakt niet beviel.

BESTAND: 1.24 MQM Kimi.json

1. "afgrijst"

Categorie : ACCURACY > Mistranslation

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Zie jij die man, Decianus, met zijn ongekamde haren,
van wiens strenge voorhoofd zelfs jij je afgrijst,
die spreekt over de Curii en de Camilli als verdedigers van de deugd?
Vertrouw dat voorhoofd niet: gisteren was hij bruid.

BESTAND: 1.28 MQM Kimi.json

1. " ruikt"

Categorie : ACCURACY > Undertranslation

Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

"Wie denkt dat Acerra nog naar de wijn van gisteren ruikt,
vergist zich: Acerra drinkt altijd door tot de ochtend."

BESTAND: 1.32 MQM Kimi.json

1. "Sabidi"

Categorie : ACCURACY > Untranslated

Ernstgraad: Medium (3 pts)

2. "dit alleen kan ik zeggen"

Categorie : STYLE > Unidiomatic style

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Ik hou niet van je, Sabidi, en ik kan niet zeggen waarom:
dit alleen kan ik zeggen: ik hou niet van je.

BESTAND: 1.33 MQM Kimi.json

1. "huilt Gellia niet om"
Categorie : STYLE
Ernstgraad: Minor (1 pts)
2. "die treurt werkelijk, die zonder getuige treurt."
Categorie : STYLE
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Haar overleden vader huilt Gellia niet om, wanneer zij alleen is;
als er iemand bij is, springen tranen op commando tevoorschijn.
Wie dan ook tracht geprezen te worden, Gellia, rouwt niet;
die treurt werkelijk, die zonder getuige treurt.

BESTAND: 1.34 MQM Kimi.json

1. "voortdurend"
Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)
2. "diefstallen"
Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)
3. "overspelige man"
Categorie : ACCURACY > Overtranslation
Ernstgraad: Minor (1 pts)
4. "zelf"
Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)
5. "en genoegens die ergens verborgen blijven, zijn je niet welgevallig."
Categorie : STYLE
Ernstgraad: Minor (1 pts)
6. "zelfs"

Categorie : ACCURACY > Addition
Ernstgraad: Minor (1 pts)

7. "en met gordijn en slot"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Medium (3 pts)

8. "de boog"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Major (9 pts)

9. "monumenten"

Categorie : ACCURACY > Undertranslation
Ernstgraad: Medium (3 pts)

10. "wolvijn-nachten"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Major (9 pts)

--- ORIGINELE TEKST ---

Met onbewaakte en wijdopen deuren, Lesbia,
bega je voortdurend overspel en verberg je je diefstallen niet.
Een toeschouwer bevalt je meer dan de overspelige man zelf,
en genoegens die ergens verborgen blijven, zijn je niet welgevallig.
Maar zelfs een hoer jaagt de getuige weg, en met gordijn en slot —
slechts zelden staat er een spleet open in de boog van Summemmus.
Leer tenminste van Chione of van Ias wat schaamte is:
zelfs monumenten verbergen die vuile wolvijn-nachten.
Vind je mijn oordeel soms te streng?
Ik verbied je niet geneukt te worden, Lesbia — ik verbied je betrupt te worden.

BESTAND: 2.3 MQM Kimi.json

Geen annotaties gevonden in dit bestand.

--- ORIGINELE TEKST ---

Sextus, je bent niets schuldig, niets schuldig, Sextus, dat geven wij toe. Want iemand is pas
schuldig, Sextus, als hij kan betalen.

BESTAND: 2.7 MQM Kimi.json

1. "Attale"

Categorie : ACCURACY > Untranslated

Ernstgraad: Medium (3 pts)

2. "Attale"

Categorie : ACCURACY > Untranslated

Ernstgraad: Neutral (0 pts)

3. "bemoeial"

Categorie : ACCURACY

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Je declameert fraai, je pleit fraai, Attale,
je maakt fraaie geschiedenissen, fraaie gedichten,
stelt fraai mimen samen, fraai epigrammen,
een fraaie grammaticus, een fraaie astroloog,
en zingt fraai en danst fraai, Attale,
fraai in de lierkunst, fraai in het balspel.
Al doe je niets goed, je doet toch alles fraai.
Wil je dat ik zeg wat je bent? Een groot bemoeial.

BESTAND: 2.8 MQM Kimi.json

1. "Als er dingen in dit boek u voorkomen, lezer,
hetzij te duister, hetzij te slecht Latijn,"

Categorie : STYLE > Unidiomatic style

Ernstgraad: Minor (1 pts)

2. "te leveren"

Categorie : ACCURACY > Overtranslation

Ernstgraad: Minor (1 pts)

3. "gezondigd"

Categorie : ACCURACY > Mistranslation > Overly literal

Ernstgraad: Medium (3 pts)

4. "hart"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)

5. "maakt"

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Als er dingen in dit boek u voorkomen, lezer,
hetzij te duister, hetzij te slecht Latijn,
is het niet mijn fout: de afschrijver heeft ze bedorven
terwijl hij haastte om u de verzen te leveren.
Maar als u denkt dat niet hij, maar ik gezondigd heb,
dan zal ik geloven dat u geen hart bezit.
"Toch zijn die dingen slecht." Alsof wij het overduidelijke ontkennen!
Ze zijn slecht, maar u maakt geen betere.

BESTAND: 2.11 MQM Kimi.json

1. "voorhoofd "

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Minor (1 pts)

2. "hangt om zijn zwijgend"

Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Medium (3 pts)

3. "loom "

Categorie : ACCURACY > Mistranslation > Overly literal
Ernstgraad: Medium (3 pts)

--- ORIGINELE TEKST ---

Dat je Selius met een somber voorhoofd ziet, Rufus,
dat de wandelaar de zuilengang tot laat verslijten ziet,
dat er iets treurigs hangt om zijn zwijgend, loom gezicht,
dat zijn lelijke neus bijna de grond raakt,
dat zijn rechterhand zijn borst slaat en zijn haar uitrukt:
hij beweent niet het lot van een vriend of broer,
beide zonen leven en ik bid dat ze lang leven,
zijn vrouw is gezond en de kinderen en slaven,

de pachter noch de opzichter heeft iets verduisterd.
Welk is dan de oorzaak van zijn rouw? Hij eet thuis.

BESTAND: 2.13 MQM Kimi.json

1. "Mijn advies, Sextus, is: betaal je schuldeiser maar"
Categorie : LINGUISTIC CONVENTIONS > Grammar
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Zowel de rechter als je advocaat eist geld.
Mijn advies, Sextus, is: betaal je schuldeiser maar

=====

Bijlage M: Annotaties van de drie 'letterlijkere' vertalingen door Copilot.

MQM BATCH ANNOTATION REPORT

=====

BESTAND: 1.1 MQM Copilot LETTERLIJK.json

1. "Dit"

Categorie : ACCURACY > Mistranslation
Ernstgraad: Medium (3 pts)

2. "met wat je gaf"

Categorie : LINGUISTIC CONVENTIONS > Unintelligible
Ernstgraad: Medium (3 pts)

3. "(roem)"

Categorie : BRACKETED TEXT > Necessary for translation
Ernstgraad: Medium (3 pts)

4. "levend en nog voelend"

Categorie : STYLE > Unidiomatic style
Ernstgraad: Minor (1 pts)

5. "(iets)"

Categorie : BRACKETED TEXT > Necessary for translation
Ernstgraad: Minor (1 pts)

6. "(dood)"

Categorie : BRACKETED TEXT > Unnecessary for translation
Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Dit is degene die je leest, die je zoekt,
de overal ter wereld bekende Martialis,
door zijn spitsvondige boekjes met epigrammen:
aan wie jij, ijverige lezer, met wat je gaf
(roem) voor hem levend en nog voelend,
(iets) wat weinig dichters na hun as (dood) bezitten.

BESTAND: 1.16 MQM Copilot LETTERLIJK.json

1. "(gedichten)"

Categorie : BRACKETED TEXT > Unnecessary for translation

Ernstgraad: Neutral (0 pts)

2. "anders kan het niet"

Categorie : STYLE > Unidiomatic style

Ernstgraad: Minor (1 pts)

3. "zo is een boek nu eenmaal."

Categorie : ACCURACY > Undertranslation

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Er zijn goede (gedichten), er zijn er enkele middelmatige, er zijn er meer slechte,
 die je hier leest: anders kan het niet, Avitus—zo is een boek nu eenmaal.

BESTAND: 1.23 MQM Copilot LETTERLIJK.json

1. "de baden"

Categorie : ACCURACY > Mistranslation > Overly literal

Ernstgraad: Minor (1 pts)

--- ORIGINELE TEKST ---

Je nodigt niemand uit, Cotta, tenzij degene met wie je je wast,
 en de baden alleen geven jou een gast aan tafel.
 Ik vroeg me af waarom je mij nooit had uitgenodigd, Cotta:
 nu weet ik het: ik heb je naakt niet bevallen.

=====