

Evaluation Report for Customization of Amazon Translate Active Custom Translation with TAUS Medical/Pharma Data

Language Pair

English → Dutch

Domain

Medical/Pharma

Introduction

Online machine translation engines provide easy access to high quality machine translations. These machine translation engines are optimized for content like news articles and social media posts that end users of these online platforms frequently translate.

Businesses often want to translate text with a different style and a specific topic. For enterprise use, online machine translation engines offer customization with existing translations that reflect the desired style and topic.

TAUS makes such customization data available via the TAUS Data Marketplace and TAUS Matching Data platforms, and now AWS Marketplace.

So that business can clearly assess the value of TAUS data, Polyglot Technology LLC was tasked to independently evaluate the quality of machine translation of Amazon Translate customized with TAUS Data (using [Amazon Translate Active Custom Translation](#)) compared to non-customized Amazon Translate.

BLEU Scores for Amazon Translate and Amazon Translate Active Custom Translation

Machine Translation Evaluation

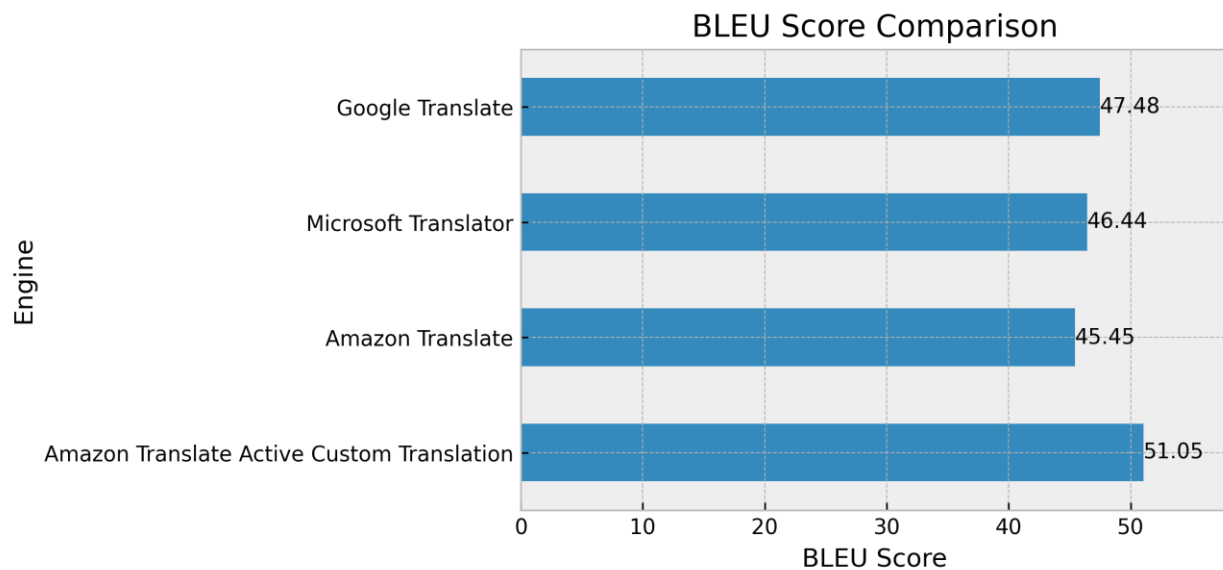
To judge whether machine translation is good or not, human evaluation is the best method. We can ask speakers of the source language and the target language, or better professional translators, to judge whether a machine translation is an adequate and fluent translation of the original text. Or we can ask how close the machine translation is to a human reference translation. Human evaluation however, is slow and hard to scale along language pairs and domains.

Automatic metrics that also use human reference translations have been developed to calculate a numeric score for machine translation quality. For close to 20 years the predominant automatic metric is BLEU, measuring the similarity of machine translations to human reference translations on a scale from 0 to 1 (or 0 to 100 when expressed as percentages). More details on BLEU and how to interpret it can be found in the section “Interpreting BLEU Scores” below.

TAUS Test Set

TAUS selects the machine translation customization data by querying its large repository of high-quality translation data with a domain-specific text. The resulting customization dataset is then split at random into a larger training set for Amazon Translate Active Custom Translation and a smaller 2,000 sentence test set that was provided to Polyglot Technology for evaluation.

BLEU Score Results for the TAUS Test Set



BLEU Score Results for Publicly Available Test Sets for the Language Pair

No widely shared public test sets are available for the language pair English-Dutch.

Description of Publicly Available Test Sets

wmt: News text test sets published by the Conference on Machine Translation¹

iwslt: Transcribed TED talk test sets published by the Conference on Spoken Language Translation²

Use Case Specific Evaluation

When employing machine translation for a specific use case, it is advisable to evaluate the systems with usage-scenario specific source text and its human reference translation. Maybe you have already data from a previous, similar project, or your translation vendor can help you create the test data. Polyglot Technology can assist in implementing a robust evaluation program.

¹ More details on the data sets can be found out e.g. for the 2020 edition on <http://statmt.org/wmt20/translation-task.html>

² Conference website: <https://iwslt.org/>

Interpreting BLEU Scores

The paragraphs in this section are adapted from Google AutoML Translate's [documentation page on evaluation](#) which is licensed under the [Creative Commons 4.0 Attribution License](#).

[BLEU \(BiLingual Evaluation Understudy\)](#) is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high quality reference translations. A value of 0 means that the machine-translated output has no overlap with the reference translation (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality).

It has been shown that BLEU scores correlate well with human judgment of translation quality. Note that even human translators do not achieve a perfect score of 1.0 (for the reason that a source sentence can have several valid, equally appropriate translations).

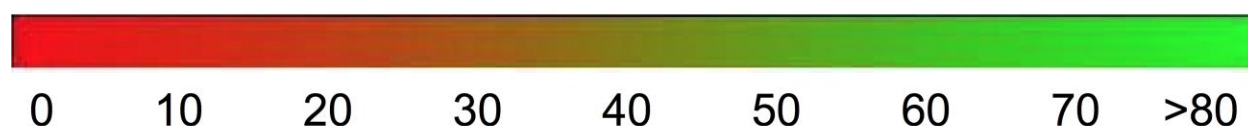
Interpretation

Trying to compare BLEU scores across different corpora and languages is strongly discouraged. Even comparing BLEU scores for the same corpus but with different numbers of reference translations can be highly misleading.

However, as a rough guideline, the following interpretation of BLEU scores (expressed as percentages rather than decimals) might be helpful.

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

The following color gradient can be used as a general scale [interpretation of the BLEU score](#):



Segment Analysis

To determine which translations improved the most between the non-customized and the customized Amazon Translate we ranked the translations by the most improved [COMET metric](#)³.

The table below shows the 30 most improved translations with a visualization of edits that have to be applied to transform the Amazon Translate translation into the Amazon Translate Active Custom Translation.

[illegible]

³ BLEU works only really well on a corpus (entire text) level than on an individual sentence level

source	reference	difference
Pertussis can also cause ear infections, bronchitis which may last a long time, pneumonia, fits, brain damage and even death.	Pertussis kan ook oorontstekingen, bronchitis die lang kan duren, longontsteking, toevallen, hersenbeschadiging en zelfs de dood veroorzaken.	Kinkhoest kan ook oorinfecties veroorzaken, bronchitis die lang kan duren, longontsteking, toevallen, hersenbeschadiging en zelfs de dood veroorzaken.
is no known antidote to fondaparinux.	Er is geen antidotum tegen fondaparinux bekend.	Er is geen bekend tegengifantidotum tegen fondaparinux bekend.
Where that is impossible for practical reasons, an enclosed leaflet, label, tape or card must contain the ingredients to which the consumer is referred either by abbreviated information or the symbol given in Annex VIII, which must appear on the container and the packaging.	Wanneer dit in de praktijk onmogelijk is, moeten de ingrediënten op een bijsluiter, een etiket, strook of kaart, gevoegd bij het product, worden vermeld; in dit geval moet de consument door middel van een verkorte aanduiding of het in bijlage VIII aangegeven symbool, die op de verpakking moeten zijn aangebracht, naar bedoelde ingrediënten worden verwezen.	Wanneer dat om praktische redenen onmogelijk is, moet en in een bijsluiter, een etiket, tapestrook of kaart de ingrediënten bevatten worden aangebracht waarnaar de consument wordt verwezen, hetzij door middel van een verkorte informatie, hetzij met het symbool in bijlage VIII aanduiding of het in bijlage VIII aangegeven symbool, dat op de verpakking container en op de verpakking moet voorkomen zijn aangebracht.
Most of the cases of tuberculosis occurred within the first eight months after initiation of therapy and may reflect recrudescence of	De meeste gevallen van tuberculose traden op in de eerste acht maanden na het starten van de therapie en kan duiden op een recidieve van een latente ziekte.	De meeste gevallen van tuberculose traden op binne in de eerste acht maanden na het beginstarten van de behandeling en kunnen een terugval van therapie en kan duiden op een recidieve
Valdoxan 25 mg also induced an advance of the time of sleep onset and of minimum heart rate.	Valdoxan 25 mg vervroegt ook de tijd van inslapen en van minimale hartfrequentie.	Valdoxan 25 mg veroorzaakte ook een voorschot van de tijd van het begin van de slaapvroegt ook de tijd van inslapen en van de minimale hartslagfrequentie.
Eligible patients for this study were to have biopsy-proven HGD in Barrett's oesophagus (BO).	De patiënten die voor deze studie uitgekozen waren, moesten HGD in Barrett's oesophagus (BO) hebben dat door biopsie aangetoond was.	De patiënten die voor deze studie in aanmerking kwamen voor dit onderzoek moesten een, moesten HGD in Barrett's oesophagus (BO) hebben dat door biopsie bewezen HGD hebben in de slokdarm van Barrett (BO) was.
The line is broken using the (optional) break parameter.	De regel wordt gebroken gebruikmakend van de (optionele) break parameter.	De regel wordt verbroken met behulp van de (optionele) break parameter voor het einde.
If, however, such deposit is made less than 15 days before the beginning of the following month, this Treaty shall not enter into force until the first day of the second month after the date of such deposit.	Indien deze nederlegging echter minder dan 15 dagen vóór het begin van de eerstvolgende maand plaatsvindt, wordt de inwerkingtreding van het Verdrag verschoven naar de eerste dag van de tweede maand volgende op die nederlegging.	Indien een dergelijke aanbetaldeze nederlegging echter minder dan 15 vijftien dagen vóór het begin van de eerstvolgende maand wordt gedaan, plaatsvindt wordt de inwerking treedt dit Verdrag pas in werking oping van het Verdrag verschoven naar de eerste dag van de tweede maand na de datum van de stortvolgende op die nederlegging.

source	reference	difference
Protaphane Protaphane should not be disposed of via wastewater or household waste.	Protaphane dient niet weggegooid te worden via het afvalwater of met huishoudelijk afval.	Protaphane Protaphane mag <u>dient</u> niet via <u>weggegooid te worden via het</u> afvalwater of <u>met</u> huishoudelijk afval worden verwijderd .
What is now threatening to happen is at complete odds with this.	Wat er nu dreigt te gebeuren is daarmee volstrekt in tegenspraak.	Wat nu dreigt te gebeuren, staat hier volledig op gespannen voet <u>tegen</u> .
Where relevant, testing should be done according to appropriate guidelines which satisfy at least the requirements for testing as included in Setac - Guidance document on regulatory testing procedures for pesticides with non-target arthropods.	Eventuele proeven moeten worden uitgevoerd volgens relevante richtsnoeren die ten minste voldoen aan de voor proeven vastgestelde eisen in Setac - Guidance document on regulatory testing procedures for pesticides with non-target arthropods.	<u>Waar relevant, Eventuele proeven</u> moeten tests worden uitgevoerd volgens <u>passende relevante</u> richt <u>snoerlij</u> nen die ten minste voldoen aan de eisen voor <u>tests</u> zoals opgenomen in Setac - Richtsnoer voor regelge <u>het uitvoeren van proeven, die zijn vastgesteld in Setac - Guidance document on regulatory test</u> ings procedures voor pesticiden <u>met geleedpotigen die niet tot het doel behoren</u> with non-target arthropods.
EMC's 2016 Global Data Protection Index1 found that most customers are behind the curve in the maturity of their data protection strategy, with 90% still evaluating or lagging.	Uit de Global Data Protection Index uit 2016 van EMC1 blijkt dat de meeste klanten zich achteraan op de curve bevinden als het gaat om de volwassenheid van hun strategie voor databescherming, en dat 90% nog aan het evalueren is of een grote achterstand heeft.	EMC's <u>De</u> Global Data Protection Index1 van EMC 2016 ontdekte dat de meeste klanten achter <u>lop de curve zitten op</u> in de loop van de looptijd van hun <u>volwassenheid van hun strategie voor</u> gegevensbescherming <u>s</u> strategie, waarbij 90% nog steeds evalueert of achterblijft.
A white opaque body and a grey opaque cap printed with a logo and "ZONEGRAN 50" in black.	Een witte opake huls en een grijze opake dop bedrukt met een logo en "ZONEGRAN 50" in het zwart.	Een witte on <u>doorzichtige body en een grijze on</u> doorzichtige pet <u>pake huls en een grijze opake dop</u> bedrukt met een logo en „ZONEGRAN 50" in het zwart.
The evaluation procedure shall include an assessment visit to the manufacturer's premises.	De beoordelingsprocedure omvat een inspectiebezoek aan de installaties van de fabrikant.	De evaluatie <u>beoordelings</u> procedure omvat een beoordeling <u>evaluatie</u> bezoek aan de gebouwen <u>installaties</u> van de fabrikant.
Exemption requiring Commission's prior approval	Vrijstelling waarvoor voorafgaande toestemming van de Commissie vereist is	Vrijstelling die <u>waarvoor</u> voorafgaande go <u>edkeu</u> r <u>stem</u> ming van de Commissie vereist <u>is</u>
Tools & Workflow & Historical Report & Process	Extra & Werkstroom & Historisch rapport & Proces	Tools <u>Extra</u> & W <u>o</u> erk <u>flow</u> stroom & g <u>h</u> istorisch rapport & p <u>ro</u> ces
Disease activity should be regularly monitored.	Ziekteactiviteit moet regelmatig gecontroleerd worden.	De <u>ziekte</u> activiteit van de ziekte <u>moedient</u> regelmatig worden <u>gecontroleerd te worden</u> .
evidence that the detoxification process does not result in harmful residues of the chemical substance used for the detoxification (as parent compound or as reaction	bewijsmateriaal waaruit blijkt dat het zuiveringsprocedé niet tot gevolg heeft dat in het gezuiverde product schadelijke residuen ontstaan van de chemische stof die voor de zuivering wordt gebruikt	bewijs <u>materiaal</u> dat het ontgift <u>zuiverings</u> proces <u>dé</u> niet leidt tot <u>gevolg heeft dat in het gezuiverde product</u> schadelijke residuen <u>ontstaan</u> van de chemische stof die wordt gebruikt voor de ontgift <u>ing</u> substantie die voor de zuivering

source	reference	difference
product) in the detoxified product;	(als moederverbinding of als reactieproduct);	<u>wordt gebruikt</u> (als moederverbinding of als reactieproduct)- <u>in het ontgifte product</u> ;
At week 76, 84% of patients re-randomised to maintenance treatment were PASI 75 responders compared with 19% of patients re-randomised to placebo (treatment withdrawal).	In week 76 was 84% van de patiënten die bij de tweede randomisatie aan de onderhoudsbehandeling waren toegewezen responder op de PASI-75, tegen 19% van de patiënten die bij de tweede randomisatie placebo kregen (staken van de therapie).	<u>In week 76 was 84% van de patiënten die opnieuw gerandomiseerd waren naar onderhoudsbehandeling PASI-75 responders vergeleken met 19% van de patiënten die opnieuw gerandomiseerd waren naar placebo (stopzetting bij de tweede randomisatie aan de onderhoudsbehandeling waren toegewezen responder op de PASI-75, tegen 19% van de patiënten die bij de tweede randomisatie placebo kregen (staken van de behandeling).</u>
The analysis of overall survival showed an improved survival in favour of the R-FC arm (p=0.0427, log-rank test), however longer follow-up is	De analyse van de totale overleving geeft een verbetering weer ten gunste van de R-FC arm (p=0,0427, log-rank test), een langere follow-up is echter nodig om deze observatie te bevestigen.	De analyse van de <u>algehele</u> totale overleving <u>toondegeeft</u> een verbeter <u>de overleving</u> ing <u>weer</u> ten gunste van de R-FC-arm (p=0,0427, log-rank-test), <u>hoeveel</u> langer de follow-up <u>ook is is</u> echter
Instructions for use and handling Liprolog cartridges are to be used with a CE marked pen as recommended in the information provided by the device manufacturer.	Instructies voor gebruik en hantering Patronen Liprolog moeten worden gebruikt met een CE-gewaarmerkte pen zoals aanbevolen in de informatie verstrekt door de fabrikant van het hulpmiddel.	Instructies voor gebruik en <u>behandeling</u> <u>van</u> Patronen Liprolog <u>cartridges</u> moeten worden gebruikt met een CE- <u>mgewaarm</u> er <u>ing</u> - <u>kte pen</u> zoals aanbevolen in de informatie <u>die verstrekt</u> door de fabrikant van het <u>apparaat is verstrekt</u> <u>hulpmiddel</u> .
This can usually be avoided if you remain inactive for at least 4 hours after taking Zerene. • There is a chance that you may experience somnambulism (sleepwalking), including eating or driving while not fully awake with no memory of the event.	Dit kan gewoonlijk vermeden worden indien u minstens 4 uur inactief blijft nadat u Zerene heeft ingenomen. • Er bestaat een kans dat u somnambulisme (slaapwandelen) ondervindt, inclusief eten of rijden terwijl u niet helemaal wakker bent en zich niets herinnert van het voorval.	<u>Dit kan meestal worden voorkomen als u te</u> <u>gewoonlijk vermeden worden indien u minstens 4 uur in</u> <u>inname van Zerene inactief blijft.</u> • <u>Er bestaat een kans dat u last krijgt van slaapwandelen (slaapwandelen)</u> <u>ctief blijft nadat u Zerene heeft ingenomen.</u> • <u>Er bestaat een kans dat u somnambulisme (slaapwandelen) ondervindt,</u> inclusief eten of rijden terwijl u niet <u>volledig</u> <u>helemaal</u> wakker bent <u>zonderen zich niets</u> herinnering <u>at</u> van het <u>evenement</u> <u>voorval</u> .
(a) the establishment and maintenance of a system which ensures that information about all suspected adverse reactions which are reported to the personnel of the company, including its representatives, is collected and collated at a single point;	a) de totstandbrenging en het beheer van een systeem dat waarborgt dat gegevens over alle vermoedelijke bijwerkingen die aan het personeel van de onderneming, waaronder het verkooppersoneel en de artsenbezoekers, worden gemeld, op één punt worden verzameld en geordend;	<u>a) de instelt</u> <u>totstandbrenging</u> en het <u>onderhoud</u> <u>beheer</u> van een systeem dat waarborgt dat <u>informatie</u> <u>gegevens</u> over alle vermoedelijke bijwerkingen die aan het personeel van de onderneming, <u>met inbegrip van</u> <u>waaronder</u> haar vertegenwoordigers, worden <u>gemel</u> <u>rapporteerd</u> , op één punt word <u>ten</u> verzameld en <u>verzamel</u> <u>geordend</u> ;

source	reference	difference
More generally, although the registered figures in all countries should exclude those already with jobs, whether in the formal or informal economy, and should include only those available and actively search ing for work, the methods of discerning whether these conditions are met differ from the confidential survey approach of the LFS, as well as varying from country to country.	Hoewel in het algemeen in elk land de registratiecijfers mensen die werken, zowel in het formele als in het grijze circuit, buiten beschouwing zouden moeten laten en alleen die personen moeten bevatten, die beschikbaar zijn en actief naar werk zoeken, verschillen de methoden om vast te stellen of aan deze voorwaarden wordt voldaan zowel van de confidentiële aanpak van de AKT, als van land tot land.	Meer in het algemeen, hoewel de geregistreerde cijfers in alle landen degenen die al een baan hebben, <u>moeten uitsluiten, hetzij</u> in de formele of informele economie, <u>moeten uitsluiten en alleen alleen degenen die</u> beschikbaar <u>zijn</u> en actief <u>vef</u> zoek <u>tochten</u> naar werk <u>zouden</u> moeten omvatten, verschillen de methoden om te bepalen of aan deze voorwaarden is voldaan, van <u>hetde</u> vertrouwelijke <u>karakter</u> , <u>enquêteonderzoeks</u> benadering van de LFS, evenals variërend van land tot land.
Premiums or discounts arising on issued and purchased securities shall be calculated and presented as part of interest income and shall be amortised over the remaining life of the securities , either according to the straight-line method or the internal rate of return (IRR) method .	Agio 's of disagio 's bij uitgegeven of aangekochte waardepapieren worden berekend en weergegeven als een deel van de rentebaten en worden afgeschreven over de resterende looptijd van de waardepapieren , hetzij volgens de lineaire methode , hetzij volgens de IRR-methode („internal rate of return ”) .	<u>Premies of kortingen die voortvloeien uit</u> Agio 's of disagio 's bij uitgegeven <u>en of</u> aangekochte <u>effect</u> <u>waardepapieren</u> worden berekend en <u>gepresenteerd als onderweergegeven als een</u> deel van <u>de</u> rentebaten en worden <u>geamortiseerdafgeschreven</u> over de resterende looptijd van de <u>effect</u> <u>waardepapieren</u> , hetzij volgens de lineaire methode <u>of de, hetzij volgens de IRR-methode (? internal rate of return (IRR)-methode?)</u> .
Radiopharmaceutical precursor - Not intended for direct application to patients.	Radiofarmaceutische precursor – Niet bedoeld voor rechtstreekse toediening aan patiënten.	Radiofarmaceutische <u>voorloperprecursor</u> - <u>—</u> Niet <u>bedoelstemd</u> voor <u>directhtstreekse</u> toep <u>ass</u> diening <u>op</u> aan patiënten.
However, in view of the potential risks associated with severe leukocytosis, a white blood cell count should be performed at regular intervals during filgrastim therapy.	Echter, gezien de mogelijke risico' s verbonden aan ernstige leukocytose, dient het aantal witte bloedcellen regelmatig te worden gecontroleerd tijdens behandeling met filgrastim.	<u>Gezien de mogelijkpotentiële</u> risico's die gepaard gaan met ernstige leukocytose, moet echter <u>met regelmatige tussenpozen een aantal witte bloedcellen worden uitgevoerd tijdens de behandeling met filgrastimgedurende de behandeling met filgrastim</u> regelmatig het aantal witte bloedcellen worden bepaald.
Unless other requirements have been laid down as a condition of the granting of the marketing authorization by the Community, these records shall be submitted to the Agency and Member States immediately upon request or at least every six months during the first two years following authorization and once a year for the following three years.	Tenzij voor het verlenen van een vergunning door de Gemeen schap andere voorwaarden zijn gesteld, moeten deze aantekeningen aan het Bureau en aan de Lid-Staten worden voorgelegd op eerste verzoek of ten minste om de zes maanden tijdens de eerste twee jaar na de verlening van de vergunning en ten minste eenmaal per jaar tijdens de daaropvolgende drie jaar.	<u>Tenzij andere vereisten zijn vastgesteld als voorwaarde voor dehet</u> verleningen van <u>de en</u> vergunning <u>vd</u> oor <u>het in de handel brengen door de Gemeenschap, worden deze gegevens onmiddellijk opde</u> Gemeenschap andere voorwaarden zijn gesteld, moeten deze aantekeningen aan <u>het Bureau en aan de Lid-Staten worden voorgelegd op eerste</u> verzoek of ten minste om de zes maanden <u>gedurende tijdens</u> de eerste twee jaar na de vergunning <u>en lening van de vergunning</u>

source	reference	difference
		en ten minste eenmaal per jaar bij het Agentschap eBureau en aan de Lids Staten ingediend. voor de volgende drie jaar.
When compared, the subcutaneous and intramuscular routes were shown to produce the same responses. • The vaccine has been shown to reduce clinical signs and mortality and reduce virus shedding in infected chickens. • The vaccine has been shown reduce virus shedding in infected ducks. • The vaccine is able to induce antibodies in a wide range of birds. • If the circulating avian influenza field virus has a different N component to the N1 included in the vaccine, it may be possible to differentiate between vaccinated from infected birds by using a diagnostic test to detect Neuraminidase antibodies.	Bij vergelijking bleken subcutane en intramusculaire toediening dezelfde respons te geven. • Aangetoond is dat het vaccin bij geïnfecteerde kippen klinische verschijnselen en sterfte reduceert en de virusuitscheiding vermindert. • Aangetoond is dat het vaccin de virusuitscheiding bij geïnfecteerde eenden vermindert. • Het vaccin kan bij zeer uiteenlopende vogelsoorten de productie van antistoffen opwekken. • Als het circulerende vogelgriepvirus een andere N-component bevat dan de N1-component in het vaccin, kunnen gevaccineerde en geïnfecteerde vogels van elkaar worden onderscheiden aan de hand van een diagnostische test die specifiek tegen neuraminidase gerichte antistoffen aantoon.	✓In vergelijken hiermee werd aangetoond dat de subcutane en intramusculaire routes dezelfde responacties opleverden. • Het isvaccin heeft aangetoond dat het vaccin de klinische symptomklinische verschijnselen en mortaliteit vermindert en de virusuitstootafscheiding bij geïnfecteerde kippen vermindert. • Het vaccin isheeft aangetoond dat het virus uitstootverlies bij geïnfecteerde eenden vermindert. • Het vaccin is in staat om • Als het circulerende aviaire influenza veldvirus een andere N-component bevaheeft dan de N1 die in het vaccin is opgenomenzit, kan het mogelijk zijn onderscheid te maken tussen gevaccineerde vaen geïnfecteerde vogels door middelmet behulp van een diagnostische test om die neuraminidase-antilichamen op te sporen aantoon.

Results Summary

By customizing Amazon Translate Active Custom Translation with TAUS Medical/Pharma English-Dutch training data the test set BLEU score improves by more than 5 points. This is a significant improvement in machine translation quality for this language pair and domain.

This numerical score improvement is supported by analyzing segments that improved significantly according to the semantic quality measure COMET: translations became more fluent and adhered better to the terminology used in the Medical/Pharma domain.

Polyglot Technology

Polyglot Technology LLC helps customers succeed with machine translation by ensuring that they make best use of data available to them, by assessing machine translation quality independent from MT vendors and by advising customers on how to best integrate the technology with people and processes.