



Language Is No Longer Human: RAG LLMs at News Agency EFE

Juan Manuel Ruiz



Manuel Herranz



LOC.
WORLD

#LocWorld51 Dublin

Language Is No Longer Human: RAG LLMs at News Agency EFE

Juan Manuel Ruiz García is chief technology officer at **Agencia EFE**, the world's premier Spanish-language news agency, leading the development of cutting-edge technical solutions tailored to empower its journalists and drive innovation across the entire organization. EFE's primary objective is to harness the power of technology to enhance its content contribution, production, and distribution systems, ensuring that it remains at the forefront of the rapidly evolving media landscape.



Language Is No Longer Human: RAG LLMs at News Agency EFE

Manuel Herranz is the CEO and founder of **Pangeanic**, a company specializing in NLP and translation services. He holds degrees from Manchester University and Entrepreneurship from MIT. He has an industrial background (Ford, Rolls Royce, and B.I. Corporation of Japan) and believes in language automation.

Manuel is known for his expertise in NLP processes, pattern recognition, deep neural networks, and adaptive technologies, pioneering “self-training MT” (DIY!) in 2009. An advocate for Ethical AI, he also actively supports several NGOs, including initiatives for better access to female education in the African sub-Sahel.



What is a news agency for?

- **Gathers and verifies news:** Collects and checks information from various sources to ensure accuracy and reliability.
- **Writes and edits news stories:** Creates and refines news articles, headlines, and summaries for publication.
- **Distributes news to clients:** Provides news content to newspapers, magazines, online platforms, and other media outlets.
- **Provides multimedia content:** Offers photos, videos, and audio files to complement news stories.
- **Monitors and updates news:** Continuously tracks and updates news stories as new information becomes available.

... the more languages you manage, the best



#LocWorld51 Dublin

Language Is No Longer Human: RAG LLMs at News Agency EFE

EFE:**EFE:**

★Adscribed offices

Agencia EFE



- **Agencia EFE is the largest Spanish-language** multimedia news agency. International, multiformat and **multilingual**
- **3M news items per year**. Our content reaches daily to 2,000+ media publications
- 2,000 journalists from 60 nationalities, operating 24/7 from more than 180 cities in 110 countries and with **3 editorial desks in 3 continents**: Madrid, Bogotá and Bangkok.
- **Idiomatic bureaus** in Rio (Portuguese), Delhi (English), Cairo (Arabic) and Barcelona (Catalan).
- Website contains more than **55,000 pages of news**

- **Pangeanic** is renowned in the field of Language Technology, with a specialization in **Machine Translation** and **Natural Language Processing (NLP)**. We have consistently evolved from our LSP origins to pioneer AI adoption in the language industry
- **We look at MT as a Machine Learning / Data Science challenge**
- Mission: To create efficient solutions that optimize the value and processing of language data with an emphasis on privacy
- Other NLP: Annotation, Data classification, Anonymization, automated subtitling, speech, data-for-AI, AI chatbot systems

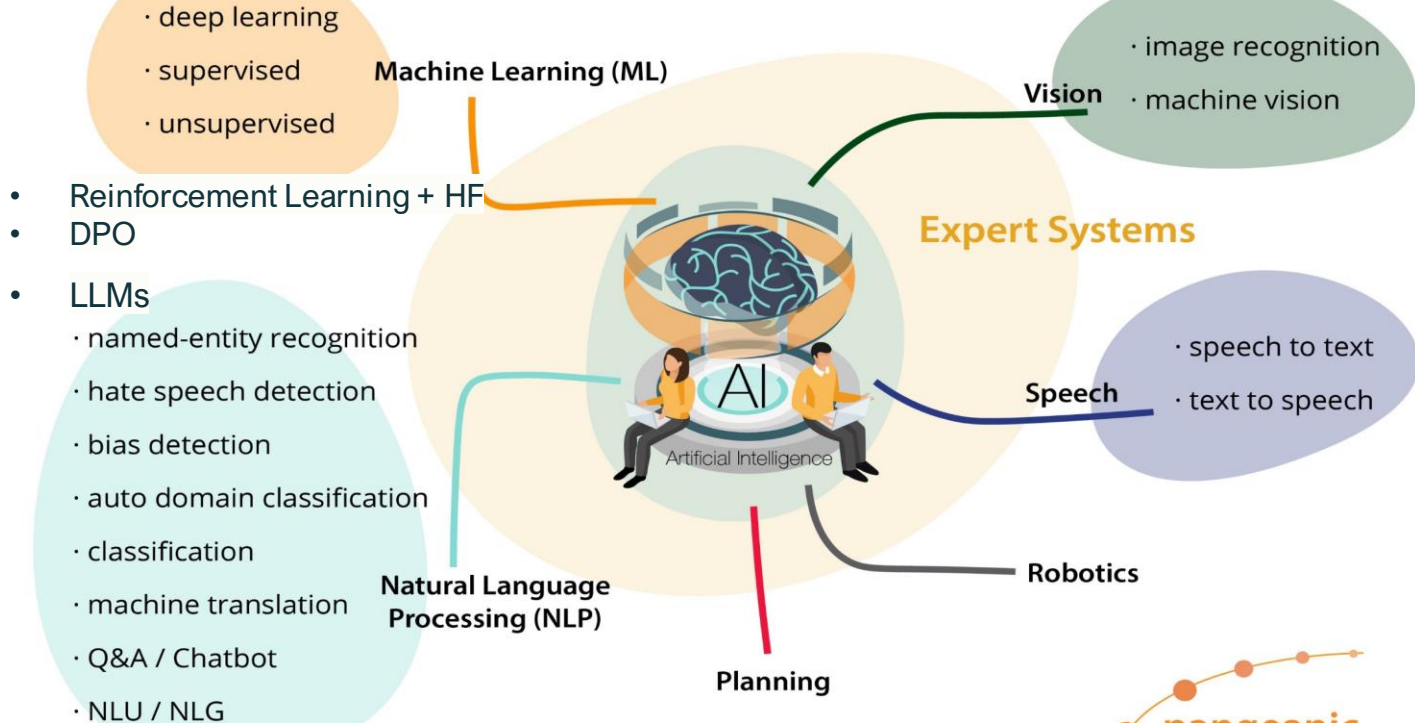


Language Is No Longer Human: RAG LLMs at News Agency EFE

Collective
Intelligence

“World”
Knowledge
(immediacy)

Language AI Disciplines



Language Is No Longer Human: RAG LLMs at News Agency EFE

Stop thinking about "language industry" and really adopt industrial, manufacturing processes.

Translation is one clear domain in which AI is excelling beyond human capabilities. [...] not easy to control LLMs' generation capabilities so that they become "true and faithful translators," [...] accuracy and terminology control can be resolved by [...] Retrieval augmented generation [...] very promising role

"AI already cracked the code of human language, no wonder language could become a commodity."

AI represents both a challenge and an opportunity for humans, fundamentally altering how we live, work, and learn. [...] imperative to strike a balance between harnessing AI's potential and safeguarding human dignity by avoiding job displacement and promoting responsible innovation — which is why I'm so vehemently against the "fair use" of copyrighted material to train models.



What truly makes us human is not language. It is the ability to respond to the unexpected. Call it "reasoning" if you wish.



#LocWorld51 Dublin



EFE in (Automatic) Translation

early 2000: first automatic translator
(Automatic Trans)

2010: Ta With You

2018: Ed. System connection (WS)

2019: Tender for Saas + API

2021: New Tender: **PANGANIC**

INTERNO

Language Is No Longer Human: RAG LLMs at News Agency EFE



How it all began...

Tender for Machine Translation in 2021.

General translation engines with a “journalism flavour” for internal use and web publishing.

Full API Management for integration in journalists’ processes.

Journalists reviews / “post-edit”.
Often divert from source to add specific or “personal” touch [*localize*].

Language Is No Longer Human: RAG LLMs at News Agency EFE

- Workflow 2021-2023... (2023 is “traditional”!!)
- Baseline > Customization > Re-training

| | |
|---|---|
| 私はパリをよく知りたい人にお勧めします。 | I would recommend it to anyone who wants to know Paris well. |
| この年の天気は暑いです。 | The weather is hot at this time of the year. |
| 私はその場所には一度も行きません。 | I would never go back to that place. |
| 東京・築地市場（中央区）の豊洲市場（江東区）への移転問題を調査する都議会の百条委員会は２０日、豊洲地区 | The 100-year committee of the metropolitan government investigating the relocation |
| 石原氏は「（都庁組織の）ピラミッドの頂点にいる人間として、（豊洲移転を）裁可したのは私の責任」と述べ、都 | Mr Ishihara said It is my responsibility to sanction (the Toyosu relocation) as a pers |
| 東京ガスに追加の土壌汚染対策費用を負担させないとする「瑕疵担保責任の免除」が、２０１１年３月の土地売買 | Regarding the circumstances in which exemption from liability for warranty against |
| おせち料理に欠かせない「かずのこ」。 | Kazunoko (herring roe) is indispensable in new year dishes. |
| 別の鍋で昆布だしをとり、醤油、みりん、日本酒を加えます。 | In another pot, boil water with dry kelp, before adding light soy sauce, mirin and sake. |
| 春になると回遊してくる何百匹ものニシンは水中で黒い塊となつてうごめき、水面に飛び跳ねるとその音が響き渡ったもの | In spring, the herring came with a roar as millions flipped at the surface and turned the water |
| ホシズナは生きているときは海藻などに付着して生活しているため、飼育にはそのような“生息場所”の提供が欠かせない。 | When alive, star sands live by attaching themselves to things like kelp, so a similar living en |
| ゲームニシンのサラダ。 | Game Salad with Salt Herring. |
| その後、漁業とは無関係の複雑な生物学的理由により、北海のニシンは数が減っているのです。 | Since then, for complex biological reasons that have nothing todo with fishing, the number c |
| 当社は、「日本をオリジンとしアジアを代表するグローバルプレイヤー」の実現に向け、世界中の市場で成長性を高めることを | In order to become a “Global Player Representing Asia with itsorigins in Japan”, Shiseido ha |
| その昔、夏は船で新鮮な魚がヘルシンキまで運ばれてきましたが、冬になると海が凍り、魚を船で運ぶことができなくなりました。 | Long ago, fresh fish was carried to Helsinki on a ship in summer, but the winter ice arrived, i |
| 企業価値の最大化を経営の最重要課題のひとつとして位置づけ、擬似カンパニー制やFCF（フリーキャッシュフロー）、DV | The Company considers the maximization of corporate value to be the most important manag daikincojp |
| ニシンのような海底産卵性の魚種は特にこの影響に対して脆弱である。 | Species of fish with demersal eggs, such as herring, are particularly vulnerable to this impac |

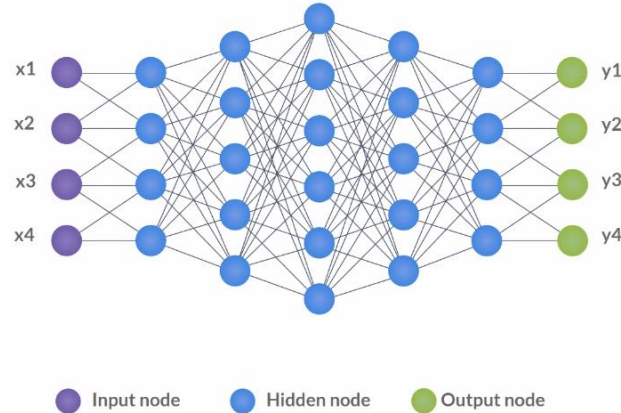
NMT



NMT neural network

Although NMT is a single neural network,, we can think of it as

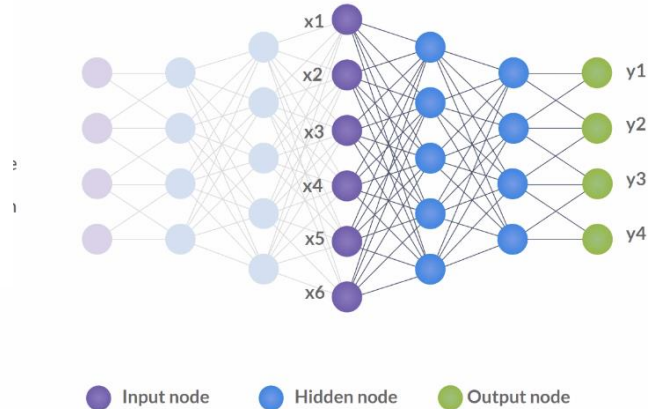
- An Encoder
- &
- A Decoder



NMT decoder

The optimized model becomes the input of the decoder.

The decoder uses the optimized model [from the source sentence] to predict the optimal target-language version.



Easy to understand

Pangeanic produces RAG-based Adaptive GenAI for AI Chatbots and next-generation machine translation processes.

NMT



PT: A Pangeanic produz GenAI adaptativa baseada em sistemas RAG para chatbots de IA e processos de tradução automática de última geração.

JA: Pangeanicは、AIチャットボットや次世代機械翻訳プロセス用のRAGシステムをベースとした適応型GenAIを製造しています。

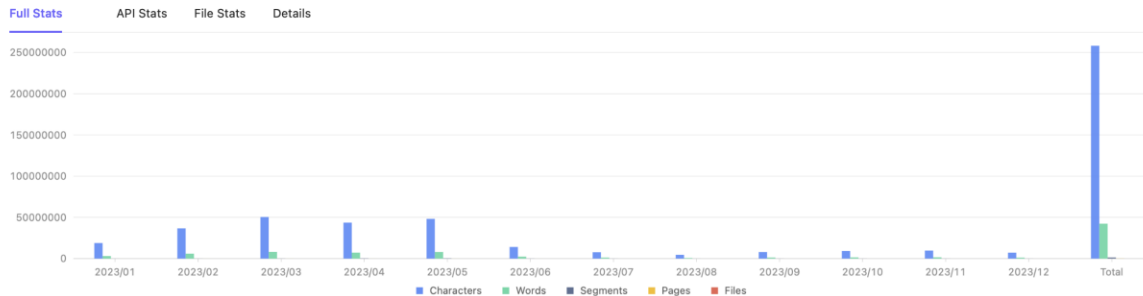
FR: Pangeanic produit une GenAI adaptative basée sur des systèmes RAG pour les chatbots d'IA et les processus de traduction automatique de nouvelle génération.

ES: Pangeanic produce GenAI adaptativa basada en sistemas RAG para chatbots con IA y procesos de traducción automática de próxima generación.

Language Is No Longer Human: RAG LLMs at News Agency EFE

2023/24

- 42,3M words machine translated in 2023
- Over 111M words since 07.2021
- Mostly web articles, internal articles for EFE's internal team of journalists
- Little NMT customization
- Journalists' preferences led to further customization in 2023



Download Csv

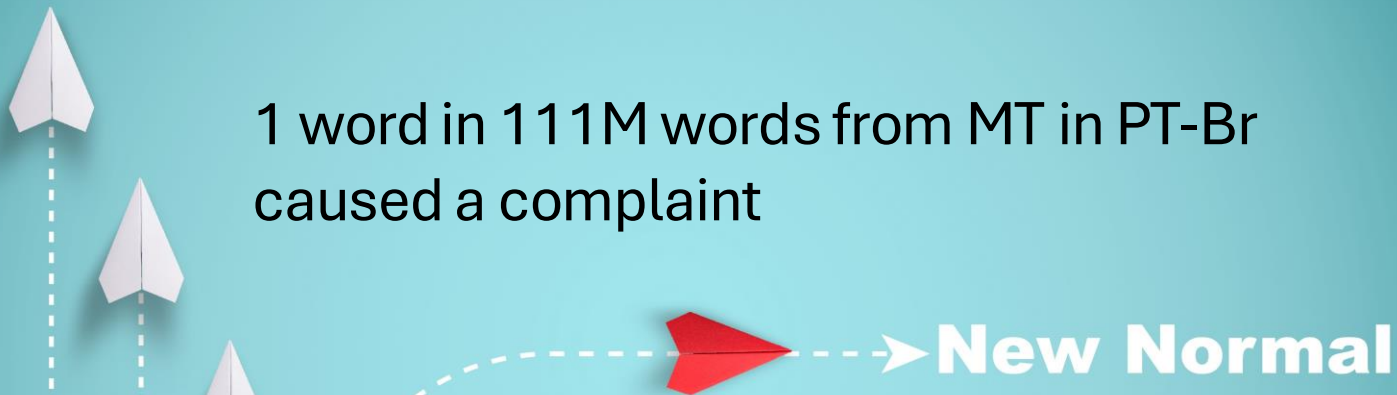
| Time | Characters | Words | Segments | Pages | Files |
|---------|-------------|------------|-----------|--------|-------|
| 2023/01 | 18.894.414 | 3.048.229 | 43.909 | 8.709 | 0 |
| 2023/02 | 36.620.078 | 5.927.766 | 78.269 | 16.936 | 0 |
| 2023/03 | 50.390.492 | 8.134.497 | 114.232 | 23.241 | 0 |
| 2023/04 | 43.572.747 | 7.216.525 | 304.676 | 20.619 | 0 |
| 2023/05 | 48.169.877 | 8.012.412 | 362.769 | 22.893 | 0 |
| 2023/06 | 14.091.676 | 2.348.315 | 107.334 | 6.709 | 0 |
| 2023/07 | 7.740.080 | 1.284.914 | 62.579 | 3.671 | 0 |
| 2023/08 | 4.598.751 | 762.273 | 38.275 | 2.178 | 0 |
| 2023/09 | 7.960.524 | 1.323.242 | 65.531 | 3.781 | 0 |
| 2023/10 | 9.180.486 | 1.520.624 | 73.828 | 4.345 | 0 |
| 2023/11 | 9.679.567 | 1.602.144 | 76.121 | 4.578 | 0 |
| 2023/12 | 7.172.642 | 1.188.174 | 57.498 | 2.296 | 0 |
| Total | 258.071.334 | 42.369.115 | 1.385.021 | 20.000 | 0 |

Talk to us

#LocWorld51 Dublin

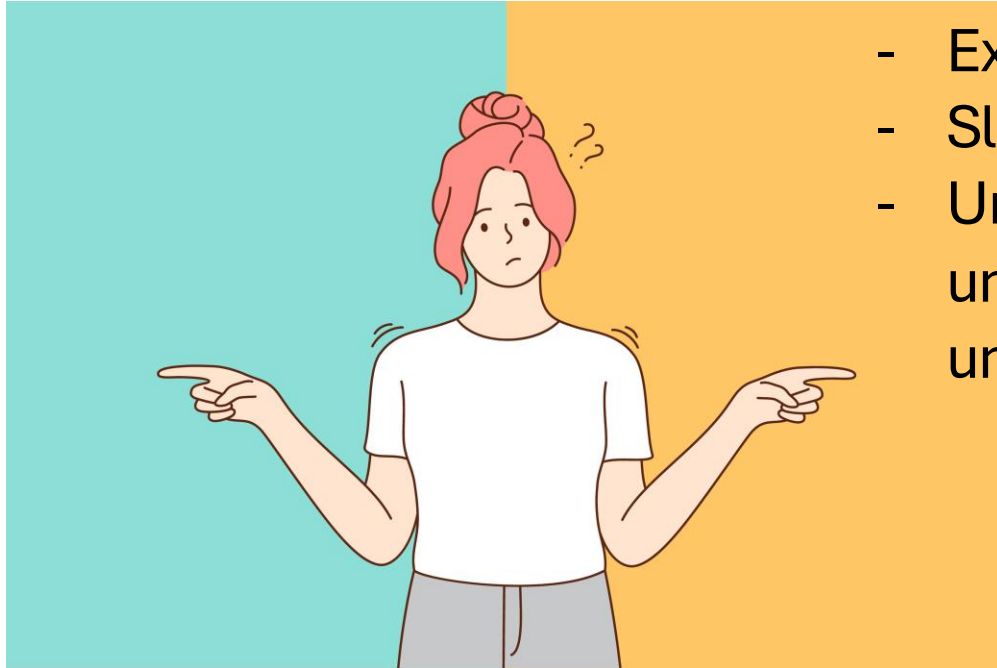
Language Is No Longer Human: RAG LLMs at News Agency EFE

1 word in 111M words from MT in PT-Br
caused a complaint



Language Is No Longer Human: RAG LLMs at News Agency EFE

- Search & clean?
- Retrain?
- Filter “stop words”



- Explore LLM?
- Slow system
- Unknown, unknown, unknown....



Language Is No Longer Human: RAG LLMs at News Agency EFE

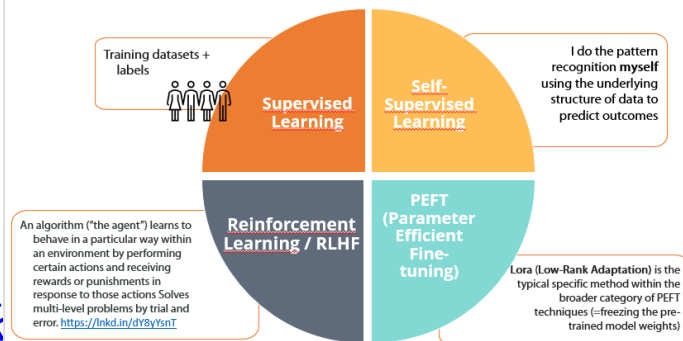
Exploiting LLMs for what they weren't designed for

2023: Dolly, Vicuna, Llama 2 (Mixtral 8x7B late '23)



- Creating a chatbot from structured data (RAG) -> Automated Post-Editing
- Entity Recognition (Anonymization/NER detection)
- MTQE – Machine Translation Quality Estimation (MT)
- Live Text/Document Classifier
- ChatGPT for MT did not beat our own NMT client engines (+ privacy issues and cost)

4 ways to fine-tune (re-train) an LLM



LoRA = Train 6% for 98% cost reduction

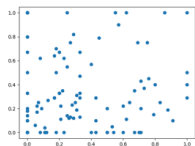
MTQE: Machine Translation Quality Estimation

RESULTS

Llama13B

- Verbosity but higher accuracy than 7B
- In 7B, Pearson correlation coefficient not significantly different than 0 (0.07961028974964038)

In 13B, we find a correlation similar to ChatGPT: 0,3!!!



RESULTS

Mixtral 8x7B

- Verbosity but higher accuracy than Llama 13B!!
- With minimum LoRA, Mixtral can "estimate" MTQE and rank in TER as good as ChatGPT 3.5
- Better than prompting: more segments processed, predictable output.
- Downstream usage

Unstructured Data Classification

OBJECTIVE

- Classify emails spam / no spam
- Llama2 7B & 13B / Mixtral vs naïve-bayes traditional method

METHODOLOGY

- 1,000 tagged emails (20 untagged = test dataset)
- Create specific prompt
- Fine-tune model
- Evaluation KPIs vs naïve-bayes

RESULTS

- Bayes: 33,76 mails/sec CPU Accuracy 0,76; Recall 0,61
- Llama 2 13B: Memory constraints, 1,45 mails/sec GPU Accuracy 0,75; Recall 0,82
- Mixtral Memory constraints, 1,5 mails/sec GPU Accuracy 0,79; Recall 0,84

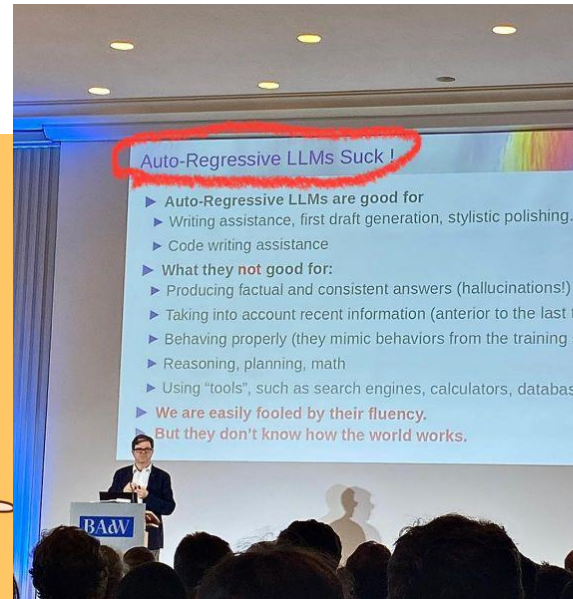
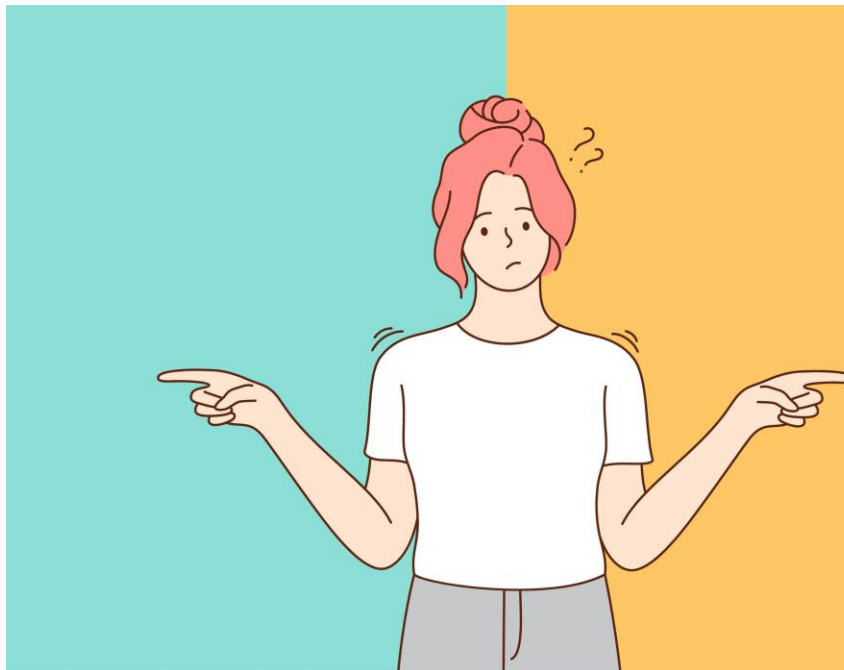
<https://pangeanic.com/use-cases/efe>

#LocWorld51 Dublin



Language Is No Longer Human: RAG LLMs at News Agency EFE

- Search & clean?
- Retrain?
- Filter “stop words”



Yann LeCun – Bavarian Academy of Science, Munich 09/29/23



Language Is No Longer Human: RAG LLMs at News Agency EFE

One LLM Project: Chatbot from structured data (RAG)

OBJECTIVE

- **In-domain Question Answering** from technical user manuals.
- Evaluate different training methodologies (dataset size, epoch, full finetune vs LoRA).

METHODOLOGY

- Automotive documentation (full website)
- H1, H2 H3 = “How to” and body is the answer
- Fine-tune model
- Evaluation KPIs

RESULTS

| | |
|---|--------------------------|
| Not useful | 35/150 (23,33%) |
| Response appears helpful, but additional text is repeated | 56/150 (37,33%) |
| Very useful, correct answer | 49/150 (32,66%) |
| No answer | 12/150 (8%) |

Off the shelf
results

Language Is No Longer Human: RAG LLMs at News Agency EFE

One LLM Project: Chatbot from structured data (RAG) -> Automated Post-Editing

OBJECTIVE

- **Search questions become TMX**
- Avoid certain expressions

METHODOLOGY

- Prioritize human input vs LLM “knowledge”
- Fine-tune model
- **Prompting**
- Verification
- Terminology
- Evaluation KPIs

RESULTS

Not useful

35/150 (23,33%)

Response very helpful, very small editing

56/150 (37,33%)

Useful, correct answer

49/150 (32,66%)

No answer

0/150 (0%)

Off the shelf
results

70% less
PE

Language Is No Longer Human: RAG LLMs at News Agency EFE

MT with RAG

version 0.0.1

Machine Translation using RAG.

Made by [Pangeanic](#).

Index name:

default

☐ Overwrite index

k

3

0

10

threshold

0.30

0.00

0.50

clear output

Upload your TM file



Drag and drop file here

Limit 200MB per file • TMX

Browse files

Translation

Source

Enter your text here

Source language

en

Target language

fr

Translate

Target

RAG Context

The context retrieved

Upload your Glossary file



Drag and drop file here

Limit 200MB per file • JSON

Browse files

Prompt

Eres un poseditor profesional de Inglés a Español. Posedita la siguiente frase, ten en cuenta las referencias proporcionadas y respeta la terminología que se usa en la REFERENCIA. Revisa que has seguido las instrucciones.

{context}

MACHINE TRANSLATION INTENT

Source: {source}

Target: {target}

AI:

Activate Windows
Go to Settings to activate Windows

#LocWorld51 Dublin



Language Is No Longer Human: RAG LLMs at News Agency EFE

Automated Metrics: BLEU Score

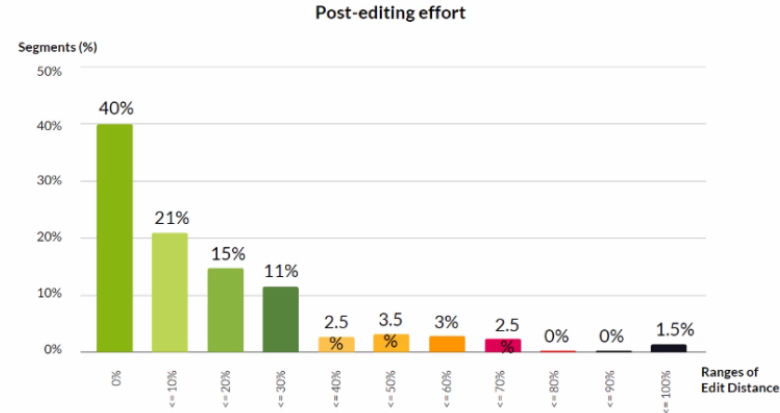
(Bilingual Evaluation Understudy)
Calculates the n-gram correspondence between the MT output for a given source sentence and a human translations of that sentence.

- Scores ranges from 0 to 1
- Test bed should be more than 1,000 sentences (and their translations) that were NOT used in the training of the system
- BLEU was designed to speed up the development of SMT.

Automated Metrics: Edit Distance

Measuring how many changes were made to MT output to create the final translation.

- The lower a score is, the better.
- Particularly helpful in measuring post-editing efficiency.
- Does not offer information on what kinds of changes were required.



(BLEU test bed: 2500 segments not included in training material)



Language Is No Longer Human: RAG LLMs at News Agency EFE

Integration in EFE's Editorial System

Editor 1 | DT3151 - FÚTBOL LIGA CAMPEONES: B.DORTMUND-REAL MADRID

Editorial interface for Editor 1. The top menu bar includes 'Navegación', 'Principal', 'Edición', and 'Herramientas'. The 'Herramientas' menu is open, showing options like 'Corrección Ortográfica', 'Sinónimos/Antónimos', 'Calendario', 'Divisas', 'RAE', 'Ver contenido original', 'Topónimos/Gentilicios', 'Diccionarios', 'Capa Lingüística', 'Traducir Documento', 'Traducir Campos', 'Calculadora', 'Web Empresarial', 'Medidas', and 'Modo Técnico'. A sidebar on the left shows language options: Español, English, Català, Português, Galego, العربية, and Française. The main workspace displays a search form with fields for 'Referencia' (DT3151), 'Relevancia', 'Estado' (Publicado), 'Clasificaciones' (Fútbol), 'Destinos' (Londres), 'Regional' (Crónica), 'Ciudad' (Londres), 'País' (Reino Unido), and 'Guía' (FÚTBOL LIGA CAMPEONES: B.DORTMUND-REAL MADRID). The article title is 'La transformación de Jadon Sancho'.

Editor 2 | DT3151 - CAMPEÕES DA FÚTBOL LIGA: B.DORTMUND X REAL MADRID

Editorial interface for Editor 2. The top menu bar includes 'Navegación', 'Principal', 'Edición', and 'Herramientas'. The 'Herramientas' menu is open, showing options like 'Nuevo', 'Crear desde', 'Cerrar Edición', 'Asignar', 'Guardar', 'Sugerencias IA', 'Publicar', 'Envío Múltiple', 'Archivar', 'Abrir Guías', 'Destinos Ordenados', 'República', 'Envío Directo', 'Despublicar', 'Desbloq', and 'Cambiar'. The main workspace displays a search form with fields for 'Referencia' (DT3151), 'Relevancia', 'Estado' (Editando), 'Clasificaciones' (Fútbol), 'Destinos' (Londres), 'Regional' (Crónica), 'Ciudad' (Londres), 'País' (Grã-Bretanha), and 'Guía' (CAMPEÕES DA FUTBOL LIGA: B.DORTMUND X REAL MADRID). The article title is 'A transformação de Jadon Sancho'.

Manuel Sánchez Gómez

Londres, 29 de maio (EFE).- Jadon Sancho viveu seus dias mais sombrios no Manchester United no ano passado. Um desentendimento com Erik Ten Hag o manteve afastado do elenco, inclusive das instalações do time principal, a ponto de não ter acesso à cantina dos companheiros de equipe e ter de levar comida para ele em uma caixa.

Tudo começou quando Ten Hag, um homem que havia demitido Cristiano Ronaldo do United meses antes, reprovou Sancho por sua atitude em uma sessão de treinamento. O holandês pediu a ele que se esforçasse mais, pois entendeu que o atacante não estava dando tudo de si quando sabia que não seria titular no jogo seguinte. Isso aconteceu a portas fechadas, mas depois que o United perdeu para o Arsenal e Sancho, como resultado desse desentendimento, não foi convocado, Ten Hag quis colocá-lo sob os olhos do público. "Em vista de seu desempenho, decidimos não convocá-lo", disse ele aos repórteres. "No United, você precisa estar em um determinado nível", acrescentou, acusando-o de estar fora de forma.

Foi um dardo que Sancho não viu com bons olhos. No dia seguinte, ele respondeu por meio da mídia social.

O ESCAPEGOAT

"Não acredite em tudo o que você lê. Não permitirei que as pessoas digam coisas que são completamente mentirosas. Eu me comportei muito bem nos treinos desta semana. Acho que há outros motivos para o que aconteceu, mas não vou entrar em detalhes. Tenho sido o bode expiatório há muito tempo e isso não é justo. Tudo o que eu quero é jogar futebol", explicou Sancho, em um claro ataque ao técnico holandês.

A resposta não foi bem recebida no United, muito menos por Ten Hag, que exigiu um pedido de desculpas, privado ou público, do jogador. O pedido de desculpas não aconteceu, embora a declaração de Sancho tenha durado apenas algumas horas em suas redes sociais antes de ser excluída.

Alguns dias depois, o United tomou uma atitude e decidiu retirar Sancho dos treinos até que uma "questão de disciplina" fosse resolvida. O jogador não teria permissão para treinar com seus companheiros de equipe, compartilhar

Language Is No Longer Human: RAG LLMs at News Agency EFE

Multiple Sources, Multiple Languages

Navegación

Mis Documentos

- ☐ AGENCIA EFE
 - ☐ CENTRAL MADRID
 - ☐ DELEGACION INTERNACIONAL
 - ☐ DELEGACION NACIONAL
 - ☒ R_AGENCIAS
 - ☒ AFP (Agence France-Presse)
 - ☒ AGI (Agenzia Giornalistica Italia)
 - ☒ ANSA Árabe (Agenzia Nazionale Stampa Associata)
 - ☒ ANSA Español (Agenzia Nazionale Stampa Associata)
 - ☒ ANSA Inglés (Agenzia Nazionale Stampa Associata)
 - ☒ ANSA Italiano (Agenzia Nazionale Stampa Associata)
 - ☒ ANSA Portugués (Agenzia Nazionale Stampa Associata)
 - ☒ BELGA (Belga News Agency)
 - ☒ Comunicados Marketwired (GlobeNewswire, Inc.)
 - ☒ LUSA (Agênciã de Notícias de Portugal, S.A)
 - ☒ PA (The Press Association)

| | | | |
|----------------------|-----------|------------------------------------|---|
| 05 jun. 2024 - 18:27 | 42991829 | SERVIÇO NACIONAL - DI-PAÍS | Homem morre atropelado em via rápida na costa sul da Madeira |
| 05 jun. 2024 - 18:27 | TX-PAR-KL | OLY-2024-FRANCE-JEUX-OLYMPIQUE | JO: les restaurateurs des quais de Seine demandent un dédommagement |
| 05 jun. 2024 - 18:26 | 0808-R01 | SPETTACOLO | === Strega: in cinquina Valerio, Di Pietrantonio, Voltolini(AGI) = |
| 8:26 | ANSA8800 | INCENDIO NELLA FONDERIA ZINKAL | Al momento non risultano feriti né persone coinvolte |
| 8:26 | ANSA8801 | ++ PUTIN, ARMI OCCIDENTALI A KIEV | 'Ancor peggio se chi le consegna le controlla anche' |
| 8:26 | ANSA8799 | ++ PUTIN, ARMI OCCIDENTALI A KIEV | 'Ancor peggio se chi le consegna le controlla anche' |
| 8:26 | HSA4426 | 1 POLITICS ELECTION 3RD LEAD | (Adds pars 10, and 24-26) |
| 8:26 | SPF000 | BRIEF/CYCLISME/ | ZLM Tour - Rune Herregodts (Intermarché-Wanty) remporte le contre-la-montre |
| 8:26 | ANSA8798 | F1: VASSEUR, FERRARI CANADÁ | |
| 8:26 | ANSA8797 | INFLUENCER E CREATOR ADERISCONI | Oggi la prima assemblea dei Content Digital Creators |
| 8:26 | ANSA8796 | PROVENZANO (PD), AGGRESSIONE E | |
| 8:25 | ANSA8795 | ANSA/PAPA: DESEMBARCO EN NORM | Mensaje al obispo de Bayeux por el 80 aniversario del Día D |
| 8:25 | 0807-R01 | ESTERO | === Strega: in cinquina Valerio, Di Pietrantonio, Voltolini(AGI) = |
| 8:25 | ANSA8794 | SUPERENALOTTO: 'SIVINCETUTTO', LI | |
| 8:25 | ANSA8793 | PATUANELLI (M5S), BEPPE C'È SEMPR | |
| 8:25 | 0806-R01 | ECONOMIA | = FOCUS = Materie prime: tornano a calare prezzi grano e mais = |
| 8:25 | ANSA8792 | CALCIO: LAMINE YAMAL PRIMO GOL | La presentazione della classifica parziale a Solomeo in Umbria |
| 8:25 | ANSA8791 | SUPERENALOTTO: 'SIVINCETUTTO', LI | |
| 8:25 | ANSA8790 | LUZI, ARMA HA SALVATO NEL 2024 G | |
| 8:24 | ANSA8789 | MIGRANTI: FOTI (FDI),ITALIA NON SA | |
| 8:24 | 0805-R01 | ESTERO | Bielorussia: Quartapelle (Pd), solidali con Bandarenka = |
| 8:24 | ANSA8787 | >>>ANSA/ WILDERS CORTEGGIA MEI | Al via il voto in Olanda e la carica dei sovranisti all'Europa |
| 05 jun. 2024 - 18:24 | ANSA8788 | SALVINI: LEGA CONTRARIA A CESSIOI | Se ci fosse un esercito europeo rischieremmo di essere in guerra |
| 05 jun. 2024 - 18:23 | ANSA8786 | EUROPEE: SERRACCHIANI, TORNIAM | 'Destra vuole smantellare la Costituzione' |
| 05 iun. 2024 - 18:23 | ANSA8785 | MFIONI MAGI? DIA77ATA CHE NON | |








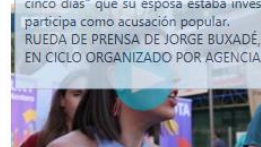












Language Is No Longer Human: RAG LLMs at News Agency EFE

Translation workflow for audiovisuals

Resultados | Pág: 1/16 | Total: 272 | Mis Documentos : AGENCIA EFE | Texto : "elecciones europeas"; Campo : Guía; Modo : Por palabra; Estado : Publicado

Texto "elecciones europeas" Campo Guía Modo Por palabra Orden Fecha de Entrada descendente Filtros Buscar

| 29 may. 2024 - 14:28 | 1368296 | 29 may. 2024 - 13:19 | 1368258 | 29 may. 2024 - 12:31 | 1368231 | 29 may. 2024 - 12:10 | 1368221 | 29 may. 2024 - 12:09 | CM3323 | 28 may. 2024 - 23:43 | 1368143 |
|--|--|---|--|--|--|------------------------------|---------|---|---------|---|---------|
| Bruto ⌚ 01:59 | Editado sin locución ⌚ 00:54 | Editado sin locución ⌚ 01:27 | Bruto ⌚ 01:57 | | | | | | | Editado sin locución ⌚ 00:41 | |
|  |  |  |  | | | | |  | |  | |
| ESP Publicado | ESP Publicado | ESP Publicado | ESP Publicado | | | | | ESP Publicado | | ESP Publicado | |
| 28 may. 2024 - 23:39 | 1368141 | | | | | | | | | | |
| Bruto ⌚ 04:09 | | | | | | | | | | | |
|  |  |  |  | | | | |  | |  | |
| ESP Publicado | ESP Publicado | ESP Publicado | ESP Publicado | | | | | ESP Publicado | | ESP Publicado | |
| 28 may. 2024 - 19:46 | CM2077 | 28 may. 2024 - 19:07 | CM2024 | 28 may. 2024 - 18:50 | CM2995 | 28 may. 2024 - 18:13 | 1368028 | 28 may. 2024 - 18:13 | 1368027 | 28 may. 2024 - 15:45 | CM2692 |
| | | | | | | Editado sin locución ⌚ 00:54 | | Bruto ⌚ 04:43 | | | |
|  |  |  |  |  |  | | | | | | |
| ESP Publicado | ESP Publicado | ESP Publicado | ESP Publicado | ESP Publicado | ESP Publicado | | | | | | |

Madrid, 29 may (EFE).- El cabeza de lista de Vox a las elecciones europeas, Jorge Buxadé, ha exigido este miércoles al presidente del Gobierno, Pedro Sánchez, explicaciones de por qué "mintió" sobre la investigación judicial de su mujer, Begoña Gómez, y le ha acusado de someter a España a un "régimen de terror".

En la segunda de las ruedas de prensa dentro del ciclo organizado por EFE con motivo de las elecciones europeas, Buxadé ha recalcado que Sánchez ya sabía cuando "se fue a descansar cinco días" que su esposa estaba investigada en la causa abierta contra ella, en la que Vox participa como acusación popular.

RUEDA DE PRENSA DE JORGE BUXADÉ, CANDIDATO DE VOX A LAS ELECCIONES EUROPEAS, EN CICLO ORGANIZADO POR AGENCIA EFE

Language Is No Longer Human: RAG LLMs at News Agency EFE

Custom MT to Custom Post-editing

Original: Los fármacos contra la obesidad han supuesto una auténtica revolución por su eficacia. Y no solo se investiga en ese campo, también en otros como en el de encontrar un arma contra la segunda enfermedad neurodegenerativa más prevalente, el párkinson. Uno de los últimos estudios publicados ha arrojado resultados con un efecto "pequeño" pero con conclusiones "interesantes".

Translation

medicamentos contra a obesidade representaram uma verdadeira revolução pela sua eficácia. E não é só nesse campo que se investiga, mas também noutros, como encontrar uma arma contra a segunda doença neurodegenerativa mais prevalente, o Parkinson. Um dos últimos estudos publicados apresentou resultados com um efeito "pequeno" mas com conclusões "interessantes".



{ "translation": "Os fármacos contra a obesidade representaram uma autêntica revolução pela sua eficácia. E não é só nesse campo que se investiga, mas também noutros, como encontrar uma arma contra a segunda doença neurodegenerativa mais prevalente, o Parkinson. Um dos últimos estudos publicados apresentou resultados com um efeito 'pequeno' mas com conclusões 'interessantes'." }



challenges

- Over post-editing (MTQE and algorithmic)
- Speed / Cost / Scale
- Prompts may vary from LLM to LLM (like MT: “not one size fits all”)
- Choice left to the end-user?

USO INTERNO

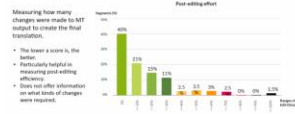
Language Is No Longer Human: RAG LLMs at News Agency EFE

Automated Metrics: BLEU Score

(Bilingual Evaluation Understudy)
Calculates the n-gram
correspondence between the MT
output for a given source
sentence and a human
translations of that sentence.

- Scores range from 0 to 1
- Test had should be more than 1,000 sentences (and their translations) that were NOT used in the training of the system
- BLEU was designed to speed up the development of SMT.

Automated Metrics: Edit Distance



(BLEU test set: 2500 segments not included in training material)

#LocWorld51 Dublin



Language Is No Longer Human: RAG LLMs at News Agency EFE



m.herranz@pangeanic.com

jruiz@efe.com

EFE:

