Dealing with hallucination and omission in neural NLG
A use case on meteorology

Jose Maria Alonso Moral, CITIUS-USC
NTP5 Invited Talk @ LORIA, Université de Lorraine, Nancy, 20 April 2022
Jose María Alonso

“Ramón y Cajal” Researcher
(RYC-2016-19802) CiTIUS-USC
Research Group on Intelligent Systems

josemaria.alonso.moral@usc.es
https://citius.usc.es/v/jose-maria-alonso-moral

https://demos.citius.usc.es/ExpliClas/
https://gitlab.citius.usc.es/jose.alonso/xai
https://gitlab.citius.usc.es/jose.alonso/bookexfs/
https://sites.google.com/view/tf-explainable-fuzzy-systems/

Task Force on Explainable Fuzzy Systems
Motivation
Data-to-text (D2T) natural language generation systems are able to process (huge) quantities of data and convert them into comprehensible texts, which contain relevant information for human users.

D2T and Fuzzy Logic – Careful Design

A use case on meteorology

- Linguistic Descriptions supported by Fuzzy Quantifiers

```
A: linguistic value
X: temperature values
Q: fuzzy quantifier

Q(∑ μₜ(x₁)/|X|)

A few days in January were warm
Between several and many days in January were cold
```

Jose M. Alonso, 20 April 2022

https://citius.usc.es/v/jose-maria-alonso-moral
Protoforms are not (in general) texts ready to be conveyed for human consumption (except for non-trivial cases)


D2T successful application

Predicción a corto plazo para Santiago de Compostela

<table>
<thead>
<tr>
<th>martes, 19 de abril</th>
<th>miércoles, 20 de abril</th>
<th>jueves, 21 de abril</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolución</td>
<td>Evolución</td>
<td>Evolución</td>
</tr>
</tbody>
</table>

Estado del cielo

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td>☀️</td>
<td>☀️</td>
<td>☁️</td>
</tr>
</tbody>
</table>

Viento

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Corriente de lluvia

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td>75%</td>
<td>75%</td>
<td>80%</td>
</tr>
</tbody>
</table>

Temperatura (°C)

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN 8°</td>
<td>MAX 16°</td>
<td>MIN 5°</td>
</tr>
</tbody>
</table>

UV Mínimo

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Avisos Meteo

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Calidad aire

<table>
<thead>
<tr>
<th>Mañana</th>
<th>Tarde</th>
<th>Noche</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comentario

Se prevén cielos con nubes y chubascos durante los próximos días. Tendremos precipitaciones todos los días. En cuanto a las temperaturas, serán normales para este periodo, con mínimas que descenderán moderadamente y máximas que se mantendrán sin cambios.

Actualización: 18/04/2022 19:00

Predicción a medio plazo para Santiago de Compostela

2 million visits per month

www.meteogalicia.gal

Jose M. Alonso, 20 April 2022

https://citius.usc.es/v/jose-maria-alonso-moral
D2T successful application

https://citius.usc.es/v/jose-maria-alonso-moral

<table>
<thead>
<tr>
<th></th>
<th>Martes, 19 de Abril</th>
<th>Miércoles, 20 de Abril</th>
<th>Jueves, 21 de Abril</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodo del día:</td>
<td>Mañana</td>
<td>Tarde</td>
<td>Noche</td>
</tr>
<tr>
<td>Estado del cielo:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viento:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperatura:</td>
<td>MIN 6 °</td>
<td>MAX 14 °</td>
<td>MIN 5 °</td>
</tr>
<tr>
<td>Calidad del aire:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comentario

Cielos parcialmente cubiertos en general durante los próximos días. Tendremos chubascos todos los días. Temperaturas normales para esta época del año, con mínimas en moderada disminución, y máximas sin cambios. En lo referido a la calidad del aire, se prevé que mejore de un grado moderado a un grado adecuado.

Jose M. Alonso, 20 April 2022

https://citius.usc.es/v/jose-maria-alonso-moral
Global Prediction: handwritten by meteorologists

WEATHERER-ES Dataset

- 3033 data records (2010 - 2020)
- Paired with linguistic descriptions in agreement with METEO's Style Guide

Input data: Table-AXYZ

Reference text: Skies will remain partly cloudy with showers, more frequent in the west.
Challenge: Global Prediction with neural NLG?
neural D2T – reusing pre-trained language models

- **Chart-to-text** (Transformer-based model)

The broadcasting revenue stream is the largest source of revenue for Liverpool FC. In 2018/2019, the football club earned approximately 299.3 million euros from domestic and international competitions broadcasting, more than twice of what they earned in 2011/2012. The second biggest revenue stream was commercial – sponsorships and merchandising.

---


After Variable Substitution:

The **Broadcasting revenue stream** is the largest source of revenue for Liverpool FC. In 2018/2019, the football club earned approximately 299.3 million euros from domestic and international competitions Broadcasting, more than twice of what they earned in 2011/2012. The second biggest **revenue stream** was Commercial – sponsorships and merchandising.

---

J. Obeid, E. Hoque, "Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model". INLG2020
[https://aclanthology.org/2020.inlg-1.20/](https://aclanthology.org/2020.inlg-1.20/)

---

Jose M. Alonso, 20 April 2022
[https://citius.usc.es/v/jose-maria-alonso-moral](https://citius.usc.es/v/jose-maria-alonso-moral)
Weather Prediction with neural D2T

- **Our approach:** (Chart-to-text-based model)
  - **WEATHER-ES Dataset**
    - Corpus in **Spanish** instead of English
      - **Min length:** 25 chars; **Max length:** 557 chars
      - **Mean length:** 186 chars; **Std length:** 71 chars
    - Data tables with categorical values which are linked by **spatio-temporal relations**
      - 3 maps per day (morning, afternoon, night)
      - 32 locations per map
      - 6 categories (19 possible values per location): Cloud, Rain, Clear, Snow, Fog, Storm
  - **NLP Pre-processing**
    - Identification of relevant tokens & named entity recognition (**verifying data-text matching**)
  - **Neural Training** (70%) & **Validation** (15%) & **Testing** (15%)
  - **NLP Post-processing**
    - Refining final texts and filling in gaps in templates with BETO model (**Spanish BERT**)
Hallucination

- “a sensory perception (such as a visual image or a sound) that occurs in the absence of an actual external stimulus and usually arises from neurological disturbance (such as that associated with delirium tremens, schizophrenia, Parkinson’s disease, or narcolepsy) or in response to drugs (such as LSD or phencyclidine)”

- delusion, illusion, hallucination, mirage mean something that is believed to be true or real but that is actually false or unreal.
  - delusion implies an inability to distinguish between what is real and what seems to be real, often as the result of a disordered state of mind.
  - illusion implies a false ascribing of reality based on what is seen or imagined.
  - hallucination implies impressions that are the product of disordered senses, because of mental illness or drugs.
  - mirage in its extended sense applies to an illusory vision, dream, hope, or aim.

We call hallucinations to texts that are incoherent or unrelated with the input of the D2T system

We call omissions to relevant pieces of information which are supported by data but not mentioned in the generated texts

Trustworthiness vs. Expectation

Faithfulness vs. Naturalness

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Quality Evaluation (input – output divergence)

- The NLP uncanny-valley Effect?

Trustworthiness vs. Expectation

Template-based NLG
+ Constrained neural NLG
Illustrative Example (input – output divergence)

**Input data:** Table-AXYZ

<table>
<thead>
<tr>
<th>Zone</th>
<th>Morning</th>
<th>Afternoon</th>
<th>Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_{01}$</td>
<td>weak showers</td>
<td>showers</td>
<td>weak showers</td>
</tr>
<tr>
<td>$Z_{02}$</td>
<td>weak showers</td>
<td>showers</td>
<td>weak showers</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$Z_{32}$</td>
<td>weak showers</td>
<td>showers</td>
<td>clouds and clearings</td>
</tr>
</tbody>
</table>

**Generated text:** The skies are expected to be cloudy with intermittent showers, occasionally stormy and accompanied by *hail*, more frequent in the morning.

**Reference text:** Skies will remain partly cloudy with showers, more frequent in the west.

Is this acceptable?

José M. Alonso, 20 April 2022

https://citius.usc.es/v/jose-maria-alonso-moral
Illustrative Example (input – output divergence)

<table>
<thead>
<tr>
<th>Zone</th>
<th>Morning</th>
<th>Afternoon</th>
<th>Night</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z_{01}</td>
<td>covered</td>
<td>clear</td>
<td>clear</td>
</tr>
<tr>
<td>Z_{02}</td>
<td>covered</td>
<td>clouds</td>
<td>clear</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Z_{02}</td>
<td>clouds</td>
<td>clear</td>
<td>high clouds</td>
</tr>
</tbody>
</table>

Generated text: Skies will be very cloudy in the morning with cloudy and clear skies and some weak rains. In the afternoon, very open skies will prevail.

Reference text: During the morning, skies will be cloudy, with more clouds on the coast. In the afternoon, skies will be partly cloudy or clear inland, while some clouds will remain on the coast.

Is this acceptable?

José M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Generated text: Los cielos estarán cubiertos, con lluvias generalizadas, más intensas y cuantiosas en el litoral atlántico, donde podrán venir acompañadas de nieve y granizo.

Hallucination sentence: Los cielos estarán cubiertos con lluvias generalizadas más intensas y cuantiosas en el litoral atlántico donde podrán venir acompañadas de nieve y granizo

Hallucination words: {'cubiertos', 'granizo', 'nieve'} --> ['litoral atlántico']

Table meteors ['litoral atlántico']: {'nublado_con_chubascos', 'lluvia'}

Is this acceptable?
Generated text: Los cielos estarán nublados por la mañana con más lluvias débiles en la franja atlántica, donde incluso habrá alguna precipitación débil. A la tarde las lluvias comenzarán por el litoral oeste y se irán extendiendo a toda Galicia.

Hallucination sentence: Los cielos estarán nublados por la mañana con más lluvias débiles en la franja atlántica donde incluso habrá alguna precipitación débil

Hallucination words: {'nublados'} --> ['franja atlántica']

Table meteors ['franja atlántica']: {'lluvia', 'lluvias débiles'}
Divergence Detector

- **Omission Detector**
  1. Look at the table with input data values -> identify all state of the sky values which apply
  2. Look at the generated text and verify if all previous values are mentioned
     - **Omission**: each value in the data table which is not mentioned in the generated text

Some omissions are appreciated because they help to yield more compact texts

**Challenge on content determination:** How to identify the most relevant pieces of information?

- **Hallucination Detector**
  1. Look at the generated text and identify all mentions related to state of the sky
  2. Look at the table with input data values -> verify if all previous mentions are supported by values in the table
     - **Hallucination**: each value which is mentioned in the generated text but it is not present in the input data

Some hallucinations are admissible and appreciated as an indicator of intelligence (they may be inferred by experts from current state of the sky)

**Challenge on content determination & serendipity:** How to infer new pieces of information which are consistent with context and common ground?
Reported Results (I)

- **Omission Detector**
  - 160 omissions out of 274 generated texts (58.39%)

- **Hallucination Detector**
  - 46 hallucinations out of 274 generated texts (16.79%)
    - 35 basic hallucinations (10 deemed as admissible by meteorologist)
    - 11 spatial hallucinations (2 deemed as admissible by meteorologist)

Further research is required to identify only omissions which are not admissible.

When contrasting the hallucinations with the corpus we observed that 4 of the cases deemed as admissible were indeed present in the corpus.

Humans also hallucinate but coherently.
Reported Results (II) – after curing the corpus with the detector

Hypothesis
- Can we improve results if we use the divergence detector to cure the corpus in advance?

Hallucination Detector
- Fixing 411 hallucinations in the training set
- 19 hallucinations out of 274 generated texts (6.93%)
  - 11 basic hallucinations
  - 8 spatial hallucinations
Related Projects & Applications
Natural Language Processing (NLP):
- Natural Language Understanding (NLU): analytics from texts
- Natural Language Generation (NLG): texts from other data sources

As a research field, NLG is in development for more than 25 years

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
2022 FIFA World Cup in Qatar
As qualifying begins to wind its way to conclusion, some teams have already booked their place in sports biggest event.

Coronavirus crisis: Patients say telehealth is OK, but most prefer to see their doctor in person.

England’s Deputy Chief Medical Officer: “Vaccines had saved around 112,000 lives and prevented more than 24 million cases of Covid.

Chinese EV maker Xpeng’s flying car company raises over $500 million as it aims to roll out in 2024.

At least 14 people killed in New York and New Jersey flooding.

The Dow is nearly 1% from its record high, while the S&P 500 and Nasdaq Composite are 1.3% and about 2.5% below their records, respectively.

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
MODELO DE CUANTIFICACIÓN

\[ Z_Q^E(A, D) = \mu_Q\left( \frac{\sum_{e \in E} \min(\mu_A(e), \mu_D(e))}{\sum_{e \in E} \mu_D(e)} \right) \]

\[ \mu_A(E') = \{0, 0, ..., 0, 1, 1, 1\} \]
Reducing, Reusing and Recycling Language Models

- BERT, RoBERTa, ALBERT, GPT-3…
  - Lack of logical and temporal reasoning or very simple deductive abilities

R. Dale, *GPT-3: What’s it good for?* Natural Language Engineering, 27(1), 113–118, 2020, [https://doi.org/10.1017/s1351324920000601](https://doi.org/10.1017/s1351324920000601)

Human: Hello, who are you? 
AI: I am an AI created by OpenAI. How can I help you today? 
Human: I am in Paris. I want a train ticket to Nancy tomorrow. 
AI: Yes, we have a train leaving at 10:30 a.m. tomorrow that will take you to Nancy. 
Human: I mean not too early, but it’s ok. Book one ticket for me, please. 
AI: Sure, I've booked a ticket for you on the 10:30 a.m. train tomorrow. Have a great day! 
Human: do you think i will have time to visite Louvre museum? 
AI: Yes, you should have plenty of time to visit the Louvre museum if you take the 10:30 a.m. train tomorrow. Thanks for asking! 
Human: Great! So, I want a ticket for tomorrow for the Louvre, what is the last ticket time that is suitable for me? 
AI: The last ticket time for the Louvre tomorrow is 9:30 p.m. Thanks for asking! 
Human: But I have to take the train at 10:30 a.m.

Jose M. Alonso, 20 April 2022
[https://citius.usc.es/v/jose-maria-alonso-moral](https://citius.usc.es/v/jose-maria-alonso-moral)
Reducing, Reusing and Recycling Language Models

- BERT, RoBERTa, ALBERT, GPT-3…
  - Lack of logical and temporal reasoning or very simple deductive abilities

R. Dale, *GPT-3: What’s it good for?* Natural Language Engineering, 27(1), 113–118, 2020, [https://doi.org/10.1017/s1351324920000601](https://doi.org/10.1017/s1351324920000601)

Human: Hello, who are you?
AI: I am an AI created by OpenAI. How can I help you today?
Human: I'd like to book an appointment with Dr Anne-Laure Rousseau.
AI: Sure, I can help you with that. Can you tell me what day and time would work for you?
Human: do you have something in the afternoon but not after 3 p.m?
AI: I have an opening at 2 p.m. tomorrow. Does that work for you?
Human: no, unfortunately
AI: How about 4 p.m.?
Human: but i'm not available after 3 p.m.
Reducing, Reusing and Recycling Language Models

- How to endow Conversational Agents with vague imprecise language and temporal reasoning based on Fuzzy Knowledge?
- How to reducing, reusing and recycling large pre-trained language models for Conversational Agents which are endowed with common-sense reasoning?

How to Ensuring Consistency while Recycling Pre-trained Language Models?

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Mobile app for prevention, cardiovascular wellbeing and rehabilitation

- Services
  - Generate textual reports about different aspects of cardiovascular health of a person

- Data
  - From external devices
  - By user
NLT + XAI => Medical Applications

- Models, techniques and methodologies based on AI for improving Medication Adherence
  - Data Exploitation Services: Symptom Patterns Discovery, Emotion Recognition
  - Persuasive Interaction: Conversational Interface with Interactive Explanations and Adaptive Communication Strategy
  - Evaluation: Clinical Trials with Polymedicated Patients; Patients with Lung Cancer; and Patients with Artrial Fibrillation

Execution Date:
2019-01-01 - 2021-12-31

Funded by
Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, Ministerio de Economía, Industria y Competitividad, RTI2018-099646-B-I00

PO FEDER Galicia 2014-2020 "Unha maneira de facer Europa"

RESEARCHERS

Main Researchers
Paulo Félix Lamas
José María Alonso Moral

CITIUS Researchers
Senén Barro Ameneiro
Jesús María Rodríguez Presedo
Purificación Carriñena Amigo
Alejandro Catalá Bolós

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Alzheimer’s Disease Diagnosis


Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
The patient is classified as AMR. In accordance with the third rule, a patient is AMR in case that the standard deviation of the number of AMR patients is big. It would be non-AMR if such standard deviation were smaller (0.345).

Final Remarks
Takeaway Messages and Open Challenges

- Content Determination remains essential in neural NLG
  - Dealing properly with **Hallucination** and **Omission** can help in the search for a good balance between **Naturalness** and **Faithfulness**

- How to deal with **Hallucination** and **Omission** in the context of ...
  - Reducing, Reusing & Recycling Pre-trained Language Models ?
  - Native generation versus post-hoc automatic Machine Translation ?
  - Low-resources Languages ?
  - Reproducibility of experiments ?
  - Responsible NLP ?
  - ACL code of Ethics ?

- What about **NLG Evaluation** ?
  - Dealing properly with **automatic metrics** and **human rating**

Jose M. Alonso, 20 April 2022
[https://citius.usc.es/v/jose-maria-alonso-moral](https://citius.usc.es/v/jose-maria-alonso-moral)
Interpretability, Explainability and Language
Be careful with oversimplification

Everything should be made as simple as possible, but not simpler.

Albert Einstein

https://citius.usc.es/v/jose-maria-alonso-moral
(*) Picture taken from the Fuzzy Archive Rudolf Selsing

From left to right: Lotfi A. Zadeh, Niels Bohr, Albert Einstein, and Werner Heisenberg

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Thank you!

Jose Maria Alonso-Moral
“Ramón y Cajal” Researcher
(RYC-2016-19802) CiTIUS-USC

josemaria.alonso.moral@usc.es
https://citius.usc.es/v/jose-maria-alonso-moral
https://gitlab.citius.usc.es/jose.alonso/xai

Task Force on Explainable Fuzzy Systems
https://sites.google.com/view/tf-explainable-fuzzy-systems/

https://sci2s.ugr.es/TF-FSS
References


E. Mariotti, Jose M. Alonso, A. Gatt. “Prometheus: Harnessing Fuzzy Logic and Natural Language for Human-centric Explainable Artificial Intelligence”. XX Spanish Congress on Fuzzy Logic and Technologies, CEDI-CAEPIA 2021

Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Joint EurAI Advanced Course on AI, TAILOR Summer School 2022

From 13/06/2022 To 17/06/2022

Organized by Intelligent Data Science and Artificial Intelligence Research Center at Universitat Politècnica de Catalunya (IDEAI-UPC)

IDEAI-UPC, Campus Nord, c. Jordi Girona, 1-3, 08034 Barcelona


Jose M. Alonso, 20 April 2022
https://citius.usc.es/v/jose-maria-alonso-moral
Supporting the right to explanation by AI
POWERED SYSTEMS

NL4XAI.eu