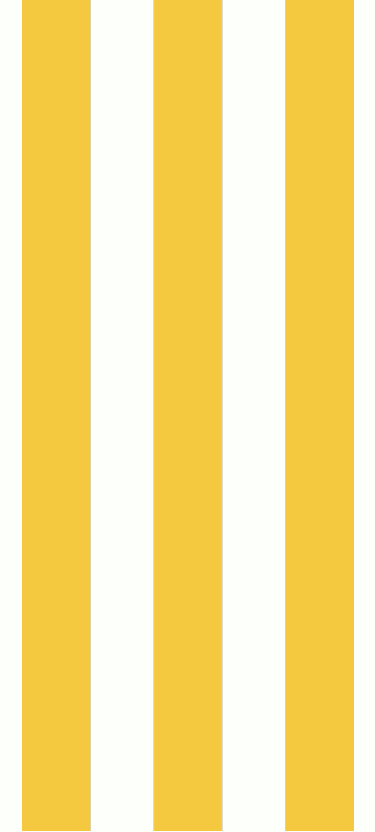


Artificial Intelligence  
(AI) for easy  
language: can AI  
simplify information?



# Speaker: Horacio Saggion

I am Horacio Saggion. Professor at Universitat Pompeu Fabra in Barcelona. I am a white male individual, I currently have a grey beard and use glasses over my green eyes. I am a computer scientist from Argentina with research interest in the area of natural language processing and artificial intelligence. In relation to this event, I have been involved for several years in projects related to inclusion and accessibility always from the view-point of language processing.



# The need for accessible content

- Universal Declaration of Human Rights recognizes access to information as a fundamental right
- The United Nations Convention on the Rights of Persons with Disabilities (CRPD) advocates for the right to accessibility
- People who encounter difficulties making meaning out of content are a diverse group of individuals



# Accessible Language

- Two methods for producing accessible communication have emerged from users and organizations concerned about the lack of accessibility of documents.
- These methods are: **easy-to-read** and **plain language**, which provide guidelines that include design and linguistic recommendations to improve the readability of printed or digital, written or oral materials.
- One of the main differences between easy-to-read and plain language is that the former emphasizes linguistic simplification and design and is based on "rules"
- Easy-to-read focuses strongly on the user who is included and involved in the creation and validation process.

# Automatic Text Simplification

- It is a *computational* process of reducing the complexity of a text by modifying the vocabulary or syntax while preserving the information and meaning.
- The goal is to make the text easier to read and understand for a wider audience.
- Example of simplification
  - *The verses were composed by the renowned bard*  
=>
  - *The famous poet wrote the poem.*

# Computational methods in text simplification

- In line with specification in easy-to-read regarding sentences and vocabulary, computational methods address tasks such as
  - Lexical simplification: such as in the previous example change of “**bard**” by “**poet**”
  - Syntactic simplification: such as in the previous example change of passive voice to the active voice
- Over the years several methods have been proposed, notably
  - **Rule-based methods** for syntactic simplification including sentence analysis and transformation
  - **Thesaurus-based methods** and psycholinguistic information (e.g. frequency, age of acquisition) for lexical simplification
  - **Discrete vectors or word imbedding** learned from corpora to model semantic relations (e.g. synonymy)
  - **Translation-based models** for sentence simplification (both syntactic and lexical)
- Many of these techniques required good quality datasets, in occasions parallel data with sentences and their correct simplifications

# Computational methods in text simplification

- Recent research in text simplification, whether syntactic or lexical, focuses primarily on the use of large language models (LLMs) and the prompting technique:
  - Natural language instructions that condition the generation of a response. For example, in the case of lexical simplification:
    - *Context: {context} Question: Given the above context, list ten alternative {language} words for "{word}" that are easier to understand. List only the words without translations, transcriptions, or explanations. Answer:*
- A considerable number of studies also use techniques that "mask" a word and extract substitutes in context using LLMs. Special attention (!) is paid to issues related to biases and prejudices in these text generation systems.
- As for production of simplified sentences, similar prompting techniques informed by easy to read guidelines can be applied.
  - For the developer eye, these LLMs perform surprisingly well, however how do we know the output is correct?

# Evaluation in text simplification

- Ideally, the end user should evaluate the text (as in the easy-to-read method).
- This is not always possible for various reasons: there is no access to the end user; the process can be costly. Often, no differences are found between the original and simplified versions because the original is already very complex.
- Automatic evaluation:
  - Comparing automatic simplification with human simplification (data is needed!)
  - Formulas used in machine translation and adapted to the problem of simplification
  - Formulas used in legibility/readability which are not always very informative

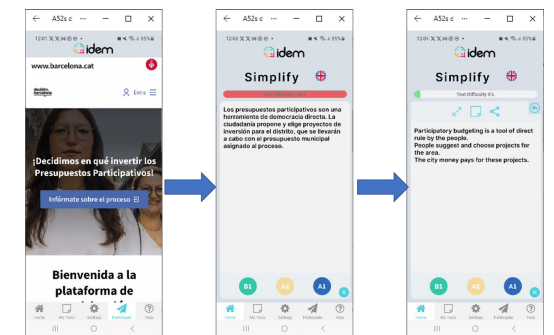
# The iDEM project: Innovative and Inclusive Democratic Spaces for Deliberation and Participation

- Several people are excluded from democratic processes because of the complexity of the language used by governments and democratic bodies
  - Electoral programs are complex
  - Information about candidates to an election is difficult to understand
  - Democratic debates are difficult to follow
  - Topics of discussion are complicated
- iDEM aims at creating spaces for deliberation and participation where information is seamlessly made easy to read and understand
  - By working with an interdisciplinary team of experts in democratic theory, user-centred design, NGOs, user organizations, NLP experts, and software integrators.



# What have we achieved so far

- Provided a better understanding of barriers (linguistic and others) for inclusion of vulnerable people in democratic processes
- Designing more inclusive democratic spaces for participation: four use cases are planned (Catalan, English, Italian, Spanish)
- Developing valuable datasets in the area of text simplification
- Developing technology for complexity assessment and transformation of text into easy to read using current techniques in NLP and AI
- Developing an app to be used in our use cases and beyond





# Can Artificial Intelligence simplify information?

- For nearly 20 years NLP has worked in the area of automatic text simplification producing valuable language resources and research.
- The question shouldn't be if AI can simplify information, it certainly can (!), but how well does it do it and how helpful the produced simplifications are for the target audience.
- In the era of Large Language Models more interesting questions remain to be answered in this field
  - Is the output appropriate?
  - Is the output faithful?
  - Is the output safe?
  - Is the output unbiased?

# Thanks for listening!

[www.idemproject.eu](http://www.idemproject.eu)

@idem\_EU 

 @idemeu.bsky.social

<https://www.linkedin.com/company/idem-project-e> 

