Using Natural Language Processing to Make the United Nations 2030 Connect Platform More Accessible

Ethan Walker Fouch, University of Virginia School of Engineering and Applied Sciences, United States of America, tyx4bc@virginia.edu

Abstract

The United Nations 2030 Connect Technological Facilitation Mechanism is an online platform created by the UN that allows access to reports and documents that address the sustainable development goals (SDGs). To make this platform more equitable for all users, the United Nations should bring together leading artificial intelligence experts to develop an AI-based language processing model to translate 2030 Connect documents into a variety of languages and convert them to various reading levels. This aligns with SDG 16.10 by creating more public access to information.

UN 2030 Connect, created in July 2020, is "an online platform that the UN Member States requested as a gateway to science, technology, and innovation resources, within and beyond the UN system, that can help the 2030 Agenda" (United Nations, n.d.-a). The 2030 Agenda is the UN's current initiative to create a more equitable and prosperous world through the implementation of 17 specific Sustainable Development Goals (United Nations, n.d.-b). The 2030 connect platform, created by the Inter-Agency Task team, helps support this effort by connecting policy makers and other members of society with the information they need to make informed decisions (IISD, n.d.). Optimizing the effectiveness of this platform and maximizing its impact is in the best interests of the UN, as this would lead to more rapid and succinct knowledge sharing.

Language Accessibility

While the UN 2030 Connect platform is a quality resource, it can be improved in terms of language accessibility by implementing technology to make the documents available in various languages and reading levels. This functionality is currently unavailable, which can exclude documents from being useful to many people around the world. Finding a solution to this issue would directly correlate with SDG 16.10, which reads "Ensure public access to information to protect fundamental freedoms, in accordance with national legislation and international agreements" (United Nations, n.d.-b). Improving the database's accessibility would ensure that the public can access important UN information and fully process and act on it. It could also set a precedent globally to have important government or research documents translated to encourage international policy collaboration on a larger scale.

Translation Disparities

A barrier to access to information that is often neglected is the language the content is written in. If an individual seeking information only speaks one language, then any documents written in other languages are useless to them. This may seem trivial, but this is the reality for many individuals from around the globe who utilize 2030 Connect. According to statistics summarized by Gitnux, over 7000 languages are spoken worldwide, with Chinese being the most spoken with 1.2 billion speakers. Additionally, the European Union (EU) recognizes 24 official languages. The EU has even seen a 32% decrease in percentage of parliament documents translated into all official languages from 2007 to 2016 (Linder, 2023). Considering the high number of languages spoken globally, and the case of the EU reducing the frequency of having its documentation translated, it is apparent that translation disparities are an issue worth addressing.

Reading Level Disparities

Reading level is a factor that is often neglected when writing a publication, partially because of the intended audience. Most publications in the technical world have the benefit of being able to utilize higher level language since their main audience is typically comprised of fellow subject matter experts. With a tool like 2030 Connect, it is important to recognize that any citizen can be a stakeholder and use the information provided, thus making reading level an important consideration. The Progress in International Reading Literacy Study shows that even in developed countries, reading levels can differ quite drastically. The study evaluates reading literacy at grade 4 and consists of countries with average index values ranging from the 300's to the low 600's, with the center point sitting at a scale of 500 (US Department of Education, n.d.). The countries scoring in the low 300's often can see this disparity carry over to their adult citizens. This can cause difficulty in everyday life as well as a basic inability to comprehend documentation as it is written. Knowing this, finding a solution to reading level disparities is paramount for the success of UN 2030 Connect.

Role of Natural Language Processing

The issues outlined previously regarding document language and reading level can be addressed by leveraging the power of Artificial Intelligence (AI), specifically Natural Language Processing (NLP). According to the National Library of Medicine, NLP is "how computers understand, process, and manipulate human languages." It often relies on semantics or patterns and involves implementation of machine learning. Chatbots and text to speech patterns are some of the most common and widely used applications of NLP (National Library of Medicine, n.d.). A language model that could translate documents or convert them to different reading levels could revolutionize the functionality of the UN 2030 Connect platform. There are a handful of issues to consider when utilizing NLP, the first of which is the capability of the technology itself. NLP Models often require diverse, high-quality data to fully learn the patterns of a language. Accuracy can also be impacted by figures of speech and other idiomatic language (Fuchs, 2023). There are also ethical concerns to consider, such as bias and privacy. NLP models can sometimes perpetuate biases that they are exposed to, and privacy can be a rising concern when used in certain applications ("Challenges and considerations in Natural Language Processing," 2024). Within the proposed application, these issues are downplayed to an extent, but it is crucial that the UN ethically implements the technology and maximizes its potential without introducing any privacy concerns.

Language Translation Technology

Using NLP for translation is a very prevalent technology. In today's world, the most common translation method is neural network translation, a variation of machine translation. This method resembles an encoder and a decoder, where a text in one language is converted to a vector and the decoder is used to decode or translate the vector into the second language. Companies such as Google and Microsoft have been using this technology since the mid 2010's, and the technology has surpassed its predecessors. Some of the modern challenges with NLP translation include uncertainty with syntax, adapting to new terminology, and the inability of language models to replicate the semantics needed. Many of the issues are caused by this last point, as some of the computational methods and statistical models can still be improved with new discoveries (Jiang and Lu, 2020). All of this suggests that with more collaboration and research, NLP translation could be further optimized to the level needed to accurately translate UN 2030 Connect documents.

Reading Level Conversion Technology

While research regarding reading level conversion is less prevalent in today's NLP landscape, the technology has lots of potential. A 2021 article from the Southern Methodist University Data Science review highlights a method used to predict reading levels of select passages using NLP. Taking in variables from different passages such as complex word count, syllable count, vocabulary size, and statistical index calculations, the model was able to successfully identify the Flesch Kincaid Grade Level (FKGL) of the passages (Arnost, Lull, Schueder, and Engler, 2021). A separate study performed and published in the International Journal of Artificial Intelligence yielded comparable results. Using a machine learning approach that considers variables such as word familiarity, word imageability, and lexical diversity, the models tested quickly identified FKGL scores to varying degrees of success. It was also found that models that consider more linguistic features tend to be more accurate (Balyan, McCarthy, and McNamara, 2020). While there has not been much research performed regarding converting one reading level into another, combining the technology that is readily used for translation as well as knowledge of the FKGL scale could lead to some breakthroughs on this front. In the meantime, simply using this reading level identification technology to label the UN documents on the FKGL scale to better inform readers could be a valuable and useful step in the right direction. This functionality could be extended into a sorting feature within the UN 2030 Connect interface that allows the user to view documents that only fall within a certain reading level or FKGL.

Policy Recommendations

NLP and its current applications indicate that the technology could be implemented to improve the UN 2030 Connect platform. The United Nations, in collaboration with the UN Commission for Science and Technology Development (CSTD), should gather leading NLP researchers from its member states to develop an NLP model for the use of translating and converting documents into the necessary languages and reading levels for use by citizens globally. While existing NLP models and technologies have seen significant growth and success, it is best to develop a new model in-house that improves on existing technologies by adding new functionality, maximizing privacy and data sovereignty, and minimizing computational costs where possible. Bringing leading experts together will allow the highest form of international collaboration, and diligent work and research can yield favorable results. If this

collaboration yields a successful language model, additional research can be performed to create a framework for accuracy and bias testing. Additionally, if any aspect of the language model proves successful in improving the 2030 Connect Platform, the UN could consider other applications for the aggregating, synthesizing, and translating of key findings, recommendations, and best practices to all member states in their preferred languages and/or reading levels. An example of this could be to introduce daily, weekly, or monthly reports produced by the UN that present essential information, news, and policy changes across the UN, with the document being available to read in a variety of languages and reading levels.

To conclude, the outcomes of this international collaboration could lead to a sustainable development revolution in the information sharing sphere, as it would establish a framework for how information should be distributed globally while enabling a wider range of recipients to be able to fully comprehend and utilize it. This could have worldwide implications for other global conglomerates such as the previously mentioned European Union. Additionally, it could lead to the education barrier having less of an impact on an individual's ability to inform themselves on current events and issues. This would be a big step forward for addressing the previously mentioned SDG 16.10 which is concerned with public access to information. On another note, it would be important for the UN to prevent this technology from making its way into national and international security sectors, as a rapid translating application could be weaponized in defense and military applications.

Acknowledgments

I would like to extend thanks to Rider Foley, my professor for Science, Technology, and Policy. His motivation and suggestions have allowed this research to surpass what I initially thought it could be. My other classmates in the class are also to be commended for their support.

References

Arnost, W., Lull, E., Schueder, J., & Engler, J. (2021). Reading Level Identification Using Natural Language Processing Techniques . SMU Data Science Review. https://scholar.smu.edu/cgi/viewcontent.cgi?article=1202&context=datasciencereview

Balyan, R., McCarthy, K. S., & McNamara, D. S. (2020). Applying Natural Language Processing and Hierarchical Machine Learning Approaches to Text Difficulty Classification. ERIC. https://eric.ed.gov/?id=ED613587

Fuchs, K. (2023, May 17). Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse?. Frontiers. https://www.frontiersin.org/articles/10.3389/feduc.2023.1166682/full

IISD. (n.d.). 2030 connect promotes science, Technology, and Innovation: News: SDG knowledge hub: IISD. SDG Knowledge Hub. https://sdg.iisd.org/news/2030-connect-promotes-science-technology-and-innovation/

Jiang, K., & Lu, X. (2020). Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review. IEEE Xplore. https://ieeexplore.ieee.org/document/9332458

Lindner, J. (2023, December 16). Must-know foreign languages statistics [current data]. GITNUX. https://gitnux.org/foreign-languages-statistics/

National Library of Medicine. (n.d.). Natural language processing. NIH National Library of Medicine. https://www.nnlm.gov/guides/data-glossary/natural-language-processing

Shelf. (2024, January 31). Challenges and considerations in Natural Language Processing. Shelf. https://shelf.io/blog/challenges-and-considerations-in-nlp/

United Nations. (n.d.-a). 2030 connect. United Nations. https://tfm2030connect.un.org/#:~:text=2030%20Connect%20is%20an%20online,help%20advance%20the%202030%20Agenda.

United Nations. (n.d.-b). Transforming our world: The 2030 agenda for sustainable development | department of economic and social affairs. United Nations. https://sdgs.un.org/2030agenda

US Department of Education. (n.d.). International Comparisons of Achievement. National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education. https://nces.ed.gov/fastfacts/display.asp?id=1