GenAI and Socially Responsible AI in Natural Language Processing Applications: A Linguistic Perspective

Christina Alexandris
National and Kapodistrian University of Athens
calexandris@gs.uoa.gr

Abstract

The evident improvement in the speed and efficiency of systems and applications with GenAI also entails some aspects that may be problematic, especially when particular text types, languages and/or user groups are concerned. The presented examples of a hybrid approach of re-introduction and integration of traditional concepts (classical theories, principles and models) may contribute to the upgrading, refinement and enrichment of training data and system output, especially in regard to the set of problematic aspects and complications in NLP systems described here, that still persist, in spite of the state-of-the-art GenAI approaches.

Problematic Aspects of State-of-the Art Data-Driven NLP Approaches

The complexity of language and its direct link to human nature and human societies is observed to be linked to complications in Natural Language Processing (NLP) that continue to persist, despite the dramatic improvement in the speed and efficiency of systems and applications with GenAI and the automated processing of vast amounts of data with state-of-the-art approaches. These complications are observed to be related to phenomena such as the adoption of human biases and higher risk of misinformation spread. However, with GenAI there are additional complications impacting socially responsible AI, especially in regard to language and its speakers and users. These complications range from special user groups to the international public and cover a broad spectrum of NLP applications involving practical tasks, from Human-Computer Interaction in services, assistance and problem-solving tasks, to (machine) translation, Information Extraction, Sentiment Analysis and Opinion Mining.

It is a widely-accepted fact that the processing of very large amounts of data with state-of-the-art Natural Language Processing (NLP) practices (i.e. Machine Learning – ML, language agnostic approaches) has resulted to a dramatic improvement in the speed and efficiency of systems and applications. However, these developments are accompanied with several challenges and difficulties that have been voiced within the last years. Characteristic examples are the challenges of spoken dialogue systems where the collaboration of experts from the Humanities is proposed (Ward and Devault 2016) and Thomas Dietterich's AAAI Presidential address referring to Machine Translation as one of the of AI applications of crucial importance calling for improvement (Dietterich 2017). We also note the ubiquitous recent articles and publications concerning the problematic aspects of applications using language models such as ChatGPT. Specifically, in regard to NLP, evident improvement in the speed and efficiency of systems and applications with GenAI also entails some aspects that may be problematic, especially when particular text types, languages and/or user groups are concerned. These aspects are:

- Aspect 1: Underrepresentation: Language Resources and natural languages – underrepresentation of less resourced languages and language groups and the speakers, cultures and nationalities they represent.
- Aspect 2: Standardization: Standardization of stylistic features and expressions, lack of diversity of expression and style.

These two aspects and their features are broadly perceived, stated and discussed. However, they are also linked to the following additional problematic aspects:

- Aspect 3: Barriers in Text Understanding (including Misunderstanding and Misinterpretation): Aspects (1) and (2) may also create barriers in understanding subtle, complex and/or implied information when text data input from less commonly used data and language resources is processed.

This also includes complex texts such as political texts and speeches, essays, manifestos, texts from historical archives, texts of religious and/or cultural significance, texts created or uttered (transcribed) by non-native speakers or by speakers of a dialect. Aspect 3 may impact mass media,Journalism, Social Media – and the outcome of human or machine decision makers alike. Furthermore, Aspect 3 may lead to the unwilling spread of misinformation.

- Aspect 4: Discouragement of HCI Usage for Special Text Types and/or User Groups: Aspects (1) and (2) may also discourage the usage and integration of customized written and spoken text output in HCI applications for special purposes and/or targeted user groups (4), especially if features such as precision, clarity, user friendliness and/or politeness are an essential requirement. This is often the case when NLP application design is heavily based on big data and, in contrast, domain-specific heuristic and rule-based approaches such as Controlled Languages (a traditional strategy for text types such as technical texts) are less commonly integrated. Aspect 4 may impact less experienced user groups and/or users who may face difficulties in interacting with a system due to circumstantial reasons (i.e. fatigue) or due to their physical or mental condition. Aspect 4 may, therefore, also contribute to the worsening of social inequalities in respect to Artificial Intelligence (AI) and AI applications.
- Aspect 5: Barriers in Accessing Information: Aspect (3) may entail that some text types and resources and their related information and knowledge may remain inaccessible to a general/ international public. As in case of Aspect 3, Aspect 5 may impact media and Journalism, Social Media– and the outcome of human or machine decision makers alike. As in Aspect 4, Aspect 5 may also contribute to the worsening of social inequalities in respect to Artificial Intelligence (AI) and AI applications.
- Aspect 6: Likelihood of Errors and False Assumptions: Aspects (3) and (4) are connected to a higher likelihood of making mistakes due to incomplete information (including false assumptions created by available data resources) and/or lack of precise and clear information in input or generated output. This applies to a wide range of written / spoken text types (i.e. from processing political-journalistic texts involving Geopolitics to medical Chatbots). Therefore, the likelihood of errors and false assumptions may also lead to the amplification of bias.
- Aspect 7: Difficulties in Error Detection and Recovery: Heavy and/or exclusive reliance on data and data-driven evaluation benchmarks may often discourage the practice of seeking and pinpointing specific types of linguistically related errors in system output for system/model improvement. This is especially common in NLP applications such as Machine Translation.

As Aspects 1 and 2 are linked to Aspects 3,4,5,6, Aspect 7 is considered to be an outcome of Aspects 3, 4, 5 and 6.

Text Types, User Requirements and Social / Individual Well-Being

In the case in Natural Language Processing applications where user requirements and data/application customization are of particular interest and importance, data may be chosen (or data input may be controlled or processed) according to application type and user groups, if necessary. Typical cases include (but are not limited to) less-resourced languages (A), less experienced users (B) and less agile users (C).

Less-resourced languages may range from natural languages that have slightly less resources than languages with the largest amounts of resources (such as English) (this is the case of most official –standard languages in the European Union) to natural languages that have very little or no resources. Natural languages that have very little or no resources are not only limited to languages spoken by isolated language communities but also include widely-spoken languages spoken by large and dynamic populations in the developing markets across the World.

Less experienced users may be a crucial factor in interactive Natural Language Processing applications such as spoken dialogue systems and ChatBots and, also, in Information Extraction and Information Retrieval. Less experienced users may require guidance by the System for an efficient interaction and/or correct choice of input. Less experienced users may include users whose profession and/or working /living environment and/or location does not involve or facilitate Human-Computer Interaction. They may also belong to particular age groups or social groups.

Less agile users may be users who face difficulties in interacting with a system or application due to factors such fatigue or mental state or due to characteristics such as age or some mild form of physical impairment. This category excludes users with severe physical or mental disabilities requiring specialized approaches, equipment and applications.

Therefore, the heavily data-driven state-of-the-art NLP approaches are proposed to include (but are not limited to) the following targets concerning social and individual well-being, according to application type:

- Making more types of information accessible to more types of recipients and user groups (i).
- Making more types of services accessible and user-friendly to more types of user groups (ii).
- Making more types of feelings, opinions, voices and reactions visible from more types of user groups (iii).

The data taken into consideration for the present user categories and targets is mostly based on a selection of registered user feedback / evaluations from projects, research papers and other sources where particular text types, languages and/or user groups are concerned. For example, elderly and/or disabled Europeans, bilingual EU residents, immigrants (European Union Projects), professional translators in the European Commission, journalists, military personnel), as well as research from the Journalism Computational Linguistics Research Lab (European Communication Institute (ECI) – Danube University of Krems (DUK), Austria).

Here, we focus on the above-described issues and present an overview of characteristic cases where the resolution of these issues can be - to a significant extent - achieved by combining state-of-the-art NLP processing approaches with classical theories, principles and models.

Re-introduction and Integration of Traditional Concepts

State-of-the-art NLP approaches with automated processing of vast amounts of data in GenAI may, indeed, also account for the above-presented cases (A,B,C) and targets (i), (ii) and (iii), in regard to the observed problematic Aspects 1-7. This can be achieved with a hybrid approach involving the re-introduction and integration of traditional concepts in state-of-the-art processing approaches, whether they are automatic or interactive. Specifically, for the above-presented issues, traditional and classical theories, principles and models are (or proposed to be) re-introduced and can be integrated into state-of-the art data-driven approaches involving Machine Learning and neural networks, functioning as training data and seed data in Natural Language Processing applications where user requirements and customization are of particular interest and importance. A hybrid approach may be considered a compromise between speed and correctness / user friendliness in (types of) NLP applications where the achievement of this balance plays a crucial role. In other words, a hybrid approach and the examples presented here target to prevent mechanisms from adopting human biases (often stemming from incomplete/lack of information, as presented above), ensuring fairness and socially responsible outcome and responsible Social Media, especially by customizing content to different linguistic and cultural groups, ensuring equitable information distribution. The way this is implemented varies according to the type of case/example/application presented, as described in the following sections. Here, we do not cover all possible cases and problems (i.e. ChatGPT) but present characteristic examples with cases employing (or benefiting from the employment of) the re-introduction of four typical types of traditional concepts concerning classical theories, principles and models. These four typical classical theories, principles and models are not considered to be flawless and without limitations, however, they can be transformed into practical strategies that can be integrated into evaluation modules, neural networks and training data (including knowledge graphs) and dialogue design. The proposed and discussed re-introduction of traditional concepts is not limited only to the particular models, principles and theories presented here. The presented examples are either applied and evaluated in applications from previous research or constitute practices from other fields and disciplines that are proposed for integration in NLP applications, with the perspective of further research and evaluations. The examples may meet all or some of the above-presented targets and user categories and may be compatible to the requirements of all or some of the Aspects 1-7, as described in the following sections.

Pragmatics – Theoretical Linguistics

From the discipline of Pragmatics in Theoretical Linguistics, the Gricean Cooperative Principle (Grice 1975) (Grice 1989) can be applied in the customization for Machine Translation and in user requirements for Spoken Dialogue Systems. In particular, the compliance to the Maxims of the Gricean Cooperative Principle contribute to the achievement of clarity and precision - important features both in Machine Translation and in spoken dialogue systems. This is especially important in (tested and evaluated) applications involving written or spoken technical texts (i.e. in the airline industry) where the traditional employment of Controlled Languages ensures the achievement of precision and clarity in system-output, with the appropriate choice of linguistic features (i.e. grammar categories, sentence types, vocabulary-terminology) (Kuhn 2014) . For example, in the case of German technical texts, participles in passive constructions that may either imply an ongoing process (Vorgangspassiv) or a finished process (Zustandspassiv) are usually avoided, due to their ambiguity (Lehrndorfer 1996).

The Gricean Cooperative Principle
Maxim of Quantity (content length and depth) Submaxims: Make your contribution as informative as is required (for current purposes of the exchange) (1). Do not make your contribution more informative than is required (2). Maxim of Quality (truth) Supermaxim: Try to make your contribution one that is true. Submaxims: Do not say what you believe is false (1). Do not say that for which you lack adequate evidence (2). Maxim of Relation (relevance)**:** Be relevant (this implies the omission of any irrelevant information). Maxim of Manner (clarity, Supermaxim: Be perspicuous. Submaxims: Avoid obscurity of expression (i.e., avoid language that is difficult to understand). (1) Avoid ambiguity (i.e., avoid language that can be interpreted in multiple ways) (2). Be brief (i.e., avoid unnecessary verbosity) (3). Be orderly- i.e. provide information in an order that makes sense, and makes it easy for the recipient to process it. (4) Source: Wikipedia

The Gricean Cooperative Principle may also help define or identify text types with multiple functions and targets (for example, to inform and to entertain) and to exclude the production of texts and or generated utterances with unwanted multiple perceived meanings and functions. This also concerns the modelling of user interaction in spoken dialog systems intended for the broad public. A typical applied and evaluated example of interaction design for banking services (Lewis 2009) involved detailed guidelines for using the appropriate expressions / sentence structure to achieve user friendliness and the system's appropriate adaptation to different user behaviors and error recovery. This is achieved by discretely encouraging user interaction and by giving the impression that the interaction is "moving forward", not "stalling" due to user / system error (Lewis 2009).

Therefore, the above-described previous research and application examples can be linked to the Gricean Cooperative Principle for overcoming barriers in text understanding (including misunderstanding and misinterpretation) (Aspect 3) and for facilitating -instead of discouraging- HCI usage for special text types and/or user groups accessing information (Aspect 4). This is achieved by making more types of information and services accessible to more types of recipients (i) and user groups (ii), namely less experienced (B) or less agile users (i.e. technical texts, banking services) (C).

On the other hand, the violation of the Maxims of the Gricean Cooperative Principle contributes to the detection and processing of Irony in written and spoken texts (Hatim 1997), especially in written and spoken political and journalistic texts and their machine translation. For example, as a violation to the Gricean Maxim of Quantity, a significantly high occurrence of adjectives, adverbials, verb-stems with very descriptive features (among other linguistic elements) was employed in NLP strategies (Machine Learning – ML) for the automatic evaluation of non-neutral content in journalistic texts from the Press (such as The Economist, Washington Post, among others) (Alexandris et. al. 2017). It should also be noted that although Irony can be derived from the context by current NLP processing practices (i.e. Machine Learning, neural networks) as presented in recent research, this is usually possible for "context-specific" cases types of Irony (i.e. Contradictory Irony or Reactionary Irony) and not for linguistic phenomena requiring an in-depth knowledge of the language(s) concerned.

The above-described previous research and application example employs the Gricean Cooperative Principle for overcoming barriers in text understanding (including misunderstanding and misinterpretation) (Aspect 3) and minimizing the likelihood of errors and false assumptions (Aspect 6), by making more types of information and services accessible to more types of recipients (i) and user groups (ii), especially less experienced users (B). In other words, these application examples employing the Gricean Cooperative Principle from the discipline of Pragmatics in Theoretical Linguistics target to contribute to the resolution of complications involving the problematic Aspects 3,4, and 6, with their compatibility to user requirements for Cases B and C and targets (i) and (ii). With its integration in system design and/or training data, the Gricean Cooperative Principle may contribute to NLP systems being more efficient and more friendly to less experienced users and/or to less agile users and to produce better quality output for text types concerning less-resourced languages and/or texts with irony and complex information.

Linguistics and Translation

From the field of Linguistics and Translation, the "Communication Pyramid" model (Desblache 2001) can be applied in the customization for Machine Translation and in user requirements for Chatbots and spoken dialogue systems. The model is proposed for integration in NLP applications, with the perspective of further research and evaluations. The higher levels of the Communication Pyramid correspond to the higher levels of knowledge and expertise of the recipients and the lower levels of the Communication Pyramid correspond to the lower levels of knowledge and expertise, with the bottom (base) of the Pyramid corresponding to the general public. The "Communication Pyramid" model can be used as a general guideline for the choice of the appropriate terminology and vocabulary and/or necessary paraphrasing according to the domain and the recipient group or user group concerned. This also includes the integration of additional, explanatory information in generated processed (i.e. translated) texts (i) or the integration of additional, explanatory information in Chatbot / Dialogue System output and/or additional steps and modules in the user's interaction with the System (ii). This approach targets to user-friendliness for a broad and varied user group in tested and evaluated applications such as banking products, as presented by researchers at IBM about a decade ago (Lewis 2009). Typical examples of integrating additional, explanatory information in generated texts (a) are the cases of the processing (i.e. Machine Translation, Information Extraction) of journalistic texts intended for a broader public with terms from Geopolitics, Medicine (as in the COVID-19 pandemic) or legislation-legal procedures (i.e. foreign citizen's rights, immigration, environmental protection, wild life conservation). Typical examples of integrating additional, explanatory information in Chatbots / Dialogue Systems are Medical Chatbots, banking services and other applications in the Service sector (b).

The level of expertise of the terminology users and the readership of the respective texts may- in some cases- determine the terminology type, linguistic parameters and nature of linguistic issues to be resolved. This also includes structure and style of the texts to be processed: languages may vary both in respect to (i) the choice of specialized terminology versus more general terms and in respect to (ii) the acceptable style and structure of texts. The perception and understanding of terminology (often originating from a foreign language) by a general public / international audience is also related to the familiarity of scientific terms and/or terms in a professional domain in the everyday life of a language community. For example, specialized financial terms become more familiar in a society where politics become increasingly connected to the International Market or where a considerable percentage of citizens are interested in the Stock Market. In this case, many financial terms may not require any explanation or reformulation.

In particular, three general categories of recipient / user types are distinguished, reflecting recipient knowledge and/or user expectations from the System (Wiegers and Beatty 2013) (Alexandris 2020). These categories are Experienced Users (existing knowledge, part of profession, culture and/or life style), Inexperienced Users (no/limited knowledge) and Distantiated Users (existing knowledge yet not part of profession, culture and/or life style) (Alexandris 2020). The choice and possible reformulation of terms within the context of sentence structure, text structure and acceptable style for the recipient / user concerned is targeted to ensure understandability, correctness, appropriateness and efficiency of the terms used.

In other words, "Communication Pyramid" model takes into account less experienced users (B) and less agile users (C), as well as experts in presenting/generating or translating information, thus, making more types of information and services accessible to more types of recipients (i) and user groups (ii). As in translation practices, applications integrating the "Communication Pyramid" model are intended to result to overcoming barriers in text understanding (misunderstanding and misinterpretation) (Aspect 3) by minimizing the likelihood of errors / false assumptions (Aspect 6). Also, and by facilitating – not discouraging - the use of applications for special text types and/or user groups in accessing information (Aspect 4) - especially from texts / text-types (and translations) that are not widely available (Aspect 5).

Therefore, the employment of the "Communication Pyramid" model from the field of Linguistics and Translation targets to contribute to the resolution of complications involving the problematic Aspects 3,4, 5 and 6, with their compatibility to user requirements for Cases B and C and targets (i) and (ii). With its integration in system design – especially in pre-editing/ post-editing modules and/or training data, the "Communication Pyramid" model may contribute to NLP systems being more efficient and friendly to less experienced users and to produce better quality output for text types concerning less-resourced languages and/or texts with specialized and/or complex information.

Linguistics and Cognitive Science

From the field of Linguistics and Cognitive Science, a set of Cognitive Bias types can be applied in the customization for Information Extraction and Information Retrieval applications, as well as for Sentiment Analysis and Opinion Mining applications. In the latter case, the identification of particular Cognitive Bias types contributes to the enrichment of training data for Sentiment Analysis and Opinion Mining applications. In particular, Lexical Bias (Trofimova 2014) concerning particular word types and the nexus of associations connected with them, is not limited to playing a crucial role in the detection of sentiment type and opinion. It may also contribute to the correct evaluation of machine translation output and the correct use of expressions and vocabulary in the output of spoken dialogue systems and ChatBots.

In Human-Computer Interaction applications, the identification of particular Cognitive Bias types also contributes to the appropriate choice of protoype-associations for linguistic and non-linguistic features (i.e. icons, images) and respective words for presenting and explaining information.

The Cognitive Biases in the user's interaction are listed as following, according to Leif Azzopardi (2021):

"Availability Bias leads people to overestimate the likelihood of an answer or stance based on how easily it can be retrieved and recalled."

"Framing Effects occur when people make different decisions given the same information because of how the information has been presented"

"Anchoring Bias stems from people's tendencies to focus too much on the first piece of information learnt, or observed (even if that information is not relevant or correct)".

"Confirmation Bias stems from people's tendency to prefer confirmatory information, where they will discount information that does not conform to their existing beliefs."

"Bandwagon Effects occur when people take on a similar opinion or point of view because other people voice that opinion or point of view".

"Researchers have been concerned that search engines may be influencing people's opinions, either by presenting confirmatory information reinforcing people's existing beliefs […], or by presenting information to sway their decisions through exposure effects (dubbed the Search Engine Manipulation Effect (SEME)." Searchers rated articles as more useful if they were easier to read and understand".

Different types of Cognitive Bias may be avoided in the interface design for the user's interaction with the System, targeting to facilitate the access to correct and complete information in Information Extraction and Information Retrieval applications. In Information Extraction (IE) applications, we note that the a-priori knowledge of expert-users may also result to Confirmation Bias, where confirmatory information may be preferred, according to one's knowledge and experience. In contrast to expert-users, non-expert users may not be aware of the types of information content in the texts concerned. Therefore, non-expert-users may be related to types of Cognitive Bias such as Availability Bias and Anchoring Bias, related to the accessibility and completeness of information.

IE and Geopolitical – Diplomatic and Military Information from Special Text Types

An example of the usage of Cognitive Bias types for the customization for Information Extraction and Information Retrieval applications is a designed user-interface and partially implemented and tested -evaluated application for accessing (diplomatic, military) knowledge from ancient classical texts. It targets to by-pass Cognitive Bias - but also to take advantage of specific types of Cognitive Bias. In particular, types of Cognitive Bias such as “Anchoring Bias”, “Confirmation Bias” and “Bandwagon Effects” (Azzopardi 2021) are avoided, whereas types of Cognitive Bias such as “Availability Bias”, and “Framing Effects” (Azzopardi 2021) are used to the advantage of the interface creation and application implementation. The main target is to allow easy access to the ancient classical texts and display detailed and/or specific information in a user-friendly interaction. Experts and professionals in the field of geopolitical and diplomatic information benefit from sources containing experience from the Past, describing geopolitical states-of-affairs, rhetoric and diplomacy, especially when complex (and not easily detectable /retrievable) information such as mentality, attitude and diplomatic skills are concerned. These resources may be a valuable yet often obscure source of information to a broader User group, requiring (a) expertise, (b) a remarkable period of time to access and to evaluate these resources in order to combine and compare information with the current state-of-affairs and (c), even, language skills.

Recent research involving approaches in extracting information from ancient texts of World History concerns two examples of ancient texts. The first case is the “Peloponnesian War” of Thucydides (Ancient Greek), taught in military academies, such as West Point, containing, among others, political and diplomatic aspects, behaviors, attitude and mentality linked to war. The second case is Sun Tzu's “The Art of War” (Ancient Chinese), summarizing the theories and principles of war, discussing practical and tactical aspects of war, the use of the special battle in war (i.e. fire attack and espionage warfare (Crawley 1903) (Venizelos 1940) (Tao 2013) (Zheng 2019) (Alexandris, Du, and Floros 2023). It should be noted that a particular structure of text and information and use of vocabulary characterizes these types of texts. This means that the employment of expert knowledge in the analysis of the text structure and content is necessary (Alexandris, Du, and Floros 2023). This is of essential importance, since the text content is not in a (dated) modern language, such as English or German - as is the case of other classical references in the domain of War (i.e. Carl Philipp Gottfried (or Gottlieb) von Clausewitz: “Vom Kriege“ (About War) (German), Alfred Thayer Mahan “The Influence of Sea Power upon History”: 1660–1783).

For the correct and efficient extraction of complex information (i.e. Diplomacy, mentalities, attitude) from the “Peloponnesian War” of Thucydides and “The Art of War” (i.e. war tactics) of Sun Tzu, the main challenge concerned was the process of guiding non-expert users and expert users alike in searching the respective information in the ancient classical texts. We note that in these cases, the search is not limited to facts, events and names and targets to access information in regard to behaviors, attitude and mentality linked to war. This is achieved with the avoidance of “Anchoring Bias”, “Confirmation Bias” and “Bandwagon Effects” (Azzopardi 2021) and with “Availability Bias”, and “Framing Effects” (Azzopardi 2021) used to the advantage of the implementation (Alexandris, Du, and Floros 2023).

The content, language and structures of these text types required a specialized customization of search techniques in addition to standard Information Extraction practices (Alexandris, Du, and Floros 2023). In particular, precision and correctness are achieved with the construction of a set of customized search ontologies combined with integrated translations by renowned scholars in languages linguistically similar to the original ancient text (capturing language-specific subtle information). The extracted passages are, subsequently connected to the respective passages in translations accessible to the international community (Alexandris, Du, and Floros 2023).

The above-described implemented application not only targets to making more types of information and services accessible to more types of recipients (i) and user groups (ii), but is also targets in making more types of feelings, opinions, voices and reactions visible from more types of user groups (iii), especially where less-resourced languages (A) are concerned. The application intends to result to overcoming barriers in text understanding (misunderstanding and misinterpretation) (Aspect 3) by minimizing the likelihood of errors / false assumptions (Aspect 6). Also, and by facilitating – not discouraging - the use of applications for special text types and/or user groups in accessing information (Aspect 4) - especially from texts / text-types (and translations) that are not widely available, such as the specialized ancient texts (Aspect 5). Therefore, these texts become accessible to a broader public, including less experienced users (B) and less agile users (C). This application example employing the Cognitive Bias concept from the field of Linguistics and Cognitive Science targets to contribute to the resolution of complications involving the problematic Aspects 3,4,5 and 6, with their compatibility to user requirements for Cases, A, B and C and targets (i), (ii) and (iii). With its integration in system design and/or seed/training data, the Cognitive Bias concept may contribute to NLP systems being more efficient and more friendly to less experienced users and/or to less agile users and to produce better quality output for text types concerning less-resourced languages and/or texts with specialized and/or complex information.

Linguistics and Psychology

From Psychology, the Plutchick Wheel Model (Plutchick 1982) can be applied in Sentiment Analysis and Opinion Mining applications as well as in the user requirements for ChatBots and spoken dialogue systems (Jurafsky and Martin 2023). The Plutchick Wheel Model can be used as a general guideline for identifying emotions and reactions in written and spoken data (including paralinguistic information). This is of particular importance for the identification of subtle emotions (i.e. "Apprehension", "Annoyance", "Disapproval", "Contempt", "Aggressiveness") (Jurafsky and Martin 2023) represented on the outer circles of the Plutchick Wheel, especially if the input originates from users with varied socio-cultural backgrounds and native languages. Furthermore, crowd-sourced input indicates that information not uttered, along with subtle emotions (occurring in the outer circles - of the Plutchik Wheel of Emotions), may be (i) differently/ falsely perceived – especially by non-native speakers of a natural language, (ii) highly dependent on random and/or circumstantial or individual-specific factors and (iii) concern specific domains and related discourse. These issues are of equal importance for Sentiment Analysis and Opinion Mining (Social Media) applications.

Integrating Implied and Unspoken Information in Knowledge Graphs

Crowd-sourced data resulted to new insights in the analysis and processing of information not uttered in spoken interaction and its integration in knowledge graphs (Alexandris, 2023), with its subsequent use in vectors and other forms of training data as dataset for training a neural network for Natural Language Processing (NLP) tasks. In the knowledge graphs generated by a (tested and evaluated) interactive application (Alexandris, Du, and Floros 2022), unspoken information is represented by the "Context" relation and the distinctive nodes it connects. The "Context" relation connects nodes with information types implied by unspoken linguistic or paralinguistic features, co-occurring with the spoken word in the utterance. In other words, the "Context" relation connected to an individual (spoken) word in a knowledge graph can shed light into the possible dimensions of the word. This also accounts for non-emotional words related to positive or negative connotations in written and spoken texts and to non-emotional words related to positive or negative input in Sentiment Analysis (and Opinion Mining) applications. The distinctive types of nodes connected by "Context" relation interaction integrated in a knowledge graph also enable a differentiation between (a) socio-culturally-biased factors/evidence and (b) circumstantial factors/evidence (individual/context-specific or domain specific -for Sentiment Analysis/HCI) in data analysis and training data) (Alexandris 2023). In the first case (a), laguage/socio-culturally-specific factors are more likely to account for speaker-participant psychology-mentality and sensitivities and for cases of intended/unintended offense or bullying, differentiating them from any random occurrences /individual-specific peculiarities (especially for paralinguistic features) and contributing to "Socially Responsible AI".

As proposed in previous research (Alexandris 2023), context-specific additional dimensions of individual spoken words integrated in knowledge graphs ("Context" relation node types (a) and (b)) may also be described as a context-specific information (atmo) "sphere" surrounding the spoken word. The concrete meaning – actual semantic content of the word (retrievable and processable in NLP) is surrounded by two context-specific layers, with its context-specific and language-specific dimensions in the inner layer of the sphere (A) and its context-specific and non-language-specific dimensions in the outer layer of the "sphere" (B) (Alexandris 2023). We also note that the outer layers of the word (atmo) "sphere" demonstrate similarities to the outer circles of the Plutchik Wheel of Emotions containing complex emotions, recognizable within a (socio-culturally determined) context, such as "contempt" and "disapproval". In contrast, concretely identifiable emotions – including intense and universally recognizable emotions, such as "rage" and "grief" - are located in the inner circles of the Plutchik Wheel of Emotions and are typically easily detected and processed by current practices in Sentiment Analysis and Opinion Mining (Alexandris 2023).

The integration of the Plutchik Wheel of Emotions targets in making more types of feelings, opinions, voices and reactions visible from more types of user groups (iii), especially where less-resourced languages (A) are concerned, but also users who are not very experienced in detecting/processing emotions, especially in a foreign language (B). The integration of the model in the knowledge graph generation application (Alexandris, Du, and Floros 2022) intends to result to overcoming barriers in text understanding (misunderstanding and misinterpretation) (Aspect 3) by minimizing the likelihood of errors / false assumptions (Aspect 6), especially from complex written/ spoken texts / text-types where emotion and/or opinion is expressed (Aspect 5). In other words, this application example (Alexandris, Du, and Floros 2022) with the employment of the Plutchick Wheel Model (Alexandris, 2023) from the discipline of Psychology targets to contribute to the resolution of complications involving the problematic Aspects 3, 5 and 6, with their compatibility to user requirements for Cases A and B and target (iii). With its integration in system design and/or seed/training data, approaches directly/indirectly linked to the Plutchick Wheel Model may contribute to NLP systems being more efficient and friendlier in assisting less experienced users in Sentiment Analysis and, also to produce better quality output for text types concerning less-resourced languages with complex information, emotion and opinion.

Conclusions and Discussion

The re-introduction and integration of traditional concepts in the data-driven state-of-the-art natural language processing approaches may contribute to the upgrading, refinement and enrichment of training data and system output. Machine Translation, ChatBots-Dialogue Systems and other Human-Computer Interaction applications can benefit from the customized integration of the Gricean Cooperation Principle and the Communication Pyramid Model, wherever necessary. The Cognitive Bias theoretical framework can play a crucial role in the correct output and efficiency of many types of applications, especially in Information Extraction. Sentiment Analysis-Opinion Mining and other Human-Computer Interaction applications involving emotion and human behavior may also benefit from the customized integration of the Plutchick Wheel Model for the detection of complex (and often unspoken) information.

The present approach and discussion mainly focused on the data preparation stage and/or the data evaluation stage in Natural Language Processing (NLP) applications where user requirements and customization are of special interest and importance. We presented a set of problematic Aspects 1-7, three cases of user requirements and three targets for overcoming complications in NLP systems that are still present in spite of the state-of-the-art approaches with GenAI. The problematic Aspects, user requirements and targets presented here may be modified or extended, according to future developments and further research.

Further research, implementation, testing-evaluation (where applicable), and any upgraded additional future implementations will confirm or contradict the efficiency of re-introducing and integrating traditional and classical theories, principles and models into Machine Learning and neural networks and other state-of-the art data-driven approaches, practices and strategies.

References

Alexandris, C.; Du, J.; and Floros, V. 2023. The Context of War and Cognitive Bias: An Interactive Approach in Accessing Relations of Attitude, Behavior and Events in Ancient Texts and Online News. *Lecture Notes in Computer Science (LNCS)* 14014: 205–223. https://doi.org/10.1007/978-3-031-35572-1_14.

Alexandris, C. 2023. Processing Information Unspoken: New Insights from Crowd-Sourced Data for Sentiment Analysis and Spoken Interaction Applications. In Proceedings of AAAI-SRAI, Socially Responsible AI for Well-being (SS-23-09) collocated with the AAAI Spring Symposium, San Francisco, CA, 2023: 90-99. https://ceur-ws.org/Vol-3527/Paper_456.pdf.

Alexandris, C.; Du, J.; and Floros, V. 2022. Visualizing and Processing Information Not Uttered in Spoken Political and Journalistic Data: From Graphical Representations to Knowledge Graphs in an Interactive Application. *Lecture Notes in Computer Science (LNCS)* 13303: 211–226. doi:10.1007/978-3-031-05409-9_16.

Alexandris, C. 2020. *Issues in Multilingual Information Processing of Spoken Political and Journalistic Texts in the Media and Broadcast News*. Newcastle upon Tyne, UK: Cambridge Scholars.

Alexandris, C.; Mylonakis, K.; Tassis, S.; Nottas, M.; and Cambourakis, G. 2017. Implementing a Platform for Complex Information Processing from Written and Spoken Journalistic Data. *Lecture Notes in Computer Science (LNCS)* 10271: 549–558.

Azzopardi, L. 2021. Cognitive Biases in Search: A review and reflection of cognitive biases in Information Retrieval. In Proceedings of the 2021 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '21). New York: Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3406522.3446023.

Crawley, R. 1903. *Thucydides' Peloponnesian War, translated by Richard Crawley*. London, UK: J.M. Dent and Co.

Desblache, L. 2001. *Aspects of Specialised Translation*. Paris, France: La Maison du Dictionnaire.

Dietterich, T. G. 2017. Steps Toward Robust Artificial Intelligence, AAAI Presidential Address. *AI Magazine* 38(3): 3-24.

Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Grice, H.P. 1975. Logic and conversation. *Syntax and Semantics* 3: 41-58.

Hatim, B. 1997. *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. Exeter, UK: University of Exeter Press.

Jurafsky, D. and Martin,J. H. 2023. *Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Preprint. https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf.

Kuhn, T. 2014. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics* 40 (1): 121-170.

Lehrndorfer, A. 1996. *Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien für eine (maschniell) kontrollierte Sprache in der technischen Dokumentation*. Tübingen, Germany: Narr.

Lewis, J.R. 2009. Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications, IBM Software Group, USA. Tutorial T09 presented at HCI 2009. San Diego, CA, July 19-24.

Plutchik, R. 1982. A psychoevolutionary theory of emotions. *Social Science Information* 21: 529-553. doi: 10.1177/053901882021004003.

Tao, Y. 2013. Translation of Verbal Anaphora in "The Art of War". *Theory and Practice in Language Studies* 3 (6): 1040-1044. doi:10.4304/tpls.3.6.1040-1044.

Trofimova, I. 2014. Observer Bias: An Interaction of Temperament Traits with Biases in the Semantic Perception of Lexical Material. *PLoSONE* 9(1): e85677.

Venizelos, E. 1940. *Thoukudidou Istoriai: Kata Metaphrasin Eleutheriou Benizelou. [Thucydides' History: translated by Eleftherios Venizelos]*. Oxford, UK: Oxford University Press.

Ward, N.G. and Devault, D. 2016. Ten Challenges in Highly-Interactive Dialog Systems. *AI Magazine* 37(4): 7-18.

Wiegers, K. and Beatty, J. 2013. *Software Requirements*. 3rd ed. London, UK: Microsoft Press, Pearson.

Zheng, J. 2019. A Comparative Study on English Translations of Military Terms in Sun Tzu: "The Art of War". *Theory and Practice in Language Studies* 9 (5): 537-544. http://dx.doi.org/10.17507/tpls.0905.07.