**ConvoKit-NLP: A Unified Approach towards Conversational Modelling Investigation**

2. Statement of the Problem: The proposal should include a statement of the problem, question or

research hypothesis to be examined by the research. Statement of the problem should include a

discussion of how the applicant came to identify this as a significant issue and the context in which

this problem emerged (for example, from the applicant’s personal experience and background, from

social science writing on the issue, or other context).

Research into Conversational Modelling is still in its infancy, blah, blah blah. Progress in Tasks such as emotion recognition/entity relation in conversation wtv idfc, release of better datasets, production of libraries to assist research, etc etc. Surely someone has tried to use it in social media algorithms. Like sure.

However, little stuff done to make application of such research feasible. Poor documentation of code, limited resources and community support, the stuff. We reason that much knowledge on the field is being overlooked when practicality is put before theory. An applied approach to research investigates the usefulness of research outside of benchmarks and what happens in theory. More attention to the field would be gained with a more comprehensive library on conversational modelling. COSMIC <https://paperswithcode.com/paper/cosmic-commonsense-knowledge-for-emotion> proves this.

3. Review of Relevant Literature: The applicant needs to make clear how the research fits into

contemporary sociology or anthropology. The review of literature should not just review general

social science literature. It should specify how the research will contribute to that literature

4. Proposed Methodology: The proposal should provide a sense of how the applicant expects to

conduct research on the problem identified. This should include a discussion of specific techniques

or methods to be used (e.g. surveys, interviews, participant observation, etc.) as well as practical

issues such as access that the applicant has to the research site, language ability requirements, etc.

We formulate our research outline into the following.

- implm TUCORE-GCN family as transformers model

- actually train a TUCORE-GCN model

- fix all the data processing bugs

- implement plug and play LLM functionality/test with some other LLMs

- put in a lot of testsuites

- write some nice documentation

- produce technical report on TUCORE-GCN

- extend to Hi-Dialog model (replace gcn with graph transformers, implement hierarchical self attention, current state of the art on some benchmarks)

- Investigate necessities of an ML developer for utilizing a model in engineering some solution.

-

5. Expected Outcomes and/or Impact of the Research: The proposal should include a discussion of

the impact of the research, both as a contribution to social science and a contribution to the people

and society where the research is to take place.

**References**

[1] Bongseok Lee and Yong Suk Choi. 2021. [Graph Based Network with Contextualized Representations of Turns in Dialogue](https://aclanthology.org/2021.emnlp-main.36). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 443–455, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[2] Xiao Liu and Jian Zhang and Heng Zhang and Fuzhao Xue and Yang You. 2023. [Hierarchical Dialogue Understanding with Special Tokens and Turn-level Attention](https://arxiv.org/abs/2305.00262).

[3] Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. [ConvoKit: A Toolkit for the Analysis of Conversations](https://aclanthology.org/2020.sigdial-1.8). In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 57–60, 1st virtual meeting. Association for Computational Linguistics.

[4] Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. [COSMIC: COmmonSense knowledge for eMotion Identification in Conversations](https://aclanthology.org/2020.findings-emnlp.224). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, Online. Association for Computational Linguistics.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805).

[6] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, & Veselin Stoyanov. (2019). [RoBERTa: A Robustly Optimized BERT Pretraining Approach](https://arxiv.org/abs/1907.11692).

[7] <https://github.com/XiaoxinHe/Awesome-Graph-LLM>

[8] Po-Wei Lin, Shang-Yu Su, and Yun-Nung Chen. 2022. [TREND: Trigger-Enhanced Relation-Extraction Network for Dialogues](https://aclanthology.org/2022.sigdial-1.58). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 623–629, Edinburgh, UK. Association for Computational Linguistics.

[9] Weishi Wang, Steven C.H. Hoi, and Shafiq Joty. 2020. [Response Selection for Multi-Party Conversations with Dynamic Topic Tracking](https://aclanthology.org/2020.emnlp-main.533). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6581–6591, Online. Association for Computational Linguistics.

[10] Zhikai Chen, Haitao Mao, Hongzhi Wen, Haoyu Han, Wei Jin, Haiyang Zhang, Hui Liu, & Jiliang Tang. 2023. Label-free Node Classification on Graphs with Large Language Models (LLMS).

[11] Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. [Dialogue-Based Relation Extraction](https://aclanthology.org/2020.acl-main.444). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics.

[12] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. [MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations](https://aclanthology.org/P19-1050). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

[13]

ConvoKit-NLP: An applied study of Conversational Modelling

Theophilus Lee  
*NUS High School of Math and Science*Singapore, Singapore  
h1910153@nushigh.edu.sg

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction (*Heading 1*)

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

# Ease of Use

## Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

## Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

# Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

[1] Bongseok Lee and Yong Suk Choi. 2021. [Graph Based Network with Contextualized Representations of Turns in Dialogue](https://aclanthology.org/2021.emnlp-main.36). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 443–455, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[2] Xiao Liu and Jian Zhang and Heng Zhang and Fuzhao Xue and Yang You. 2023. [Hierarchical Dialogue Understanding with Special Tokens and Turn-level Attention](https://arxiv.org/abs/2305.00262).

[3] Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. [ConvoKit: A Toolkit for the Analysis of Conversations](https://aclanthology.org/2020.sigdial-1.8). In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 57–60, 1st virtual meeting. Association for Computational Linguistics.

[4] Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. [COSMIC: COmmonSense knowledge for eMotion Identification in Conversations](https://aclanthology.org/2020.findings-emnlp.224). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, Online. Association for Computational Linguistics.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805).

[6] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, & Veselin Stoyanov. (2019). [RoBERTa: A Robustly Optimized BERT Pretraining Approach](https://arxiv.org/abs/1907.11692).

[7] <https://github.com/XiaoxinHe/Awesome-Graph-LLM>

[8] Po-Wei Lin, Shang-Yu Su, and Yun-Nung Chen. 2022. [TREND: Trigger-Enhanced Relation-Extraction Network for Dialogues](https://aclanthology.org/2022.sigdial-1.58). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 623–629, Edinburgh, UK. Association for Computational Linguistics.

[9] Weishi Wang, Steven C.H. Hoi, and Shafiq Joty. 2020. [Response Selection for Multi-Party Conversations with Dynamic Topic Tracking](https://aclanthology.org/2020.emnlp-main.533). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6581–6591, Online. Association for Computational Linguistics.

[10] Zhikai Chen, Haitao Mao, Hongzhi Wen, Haoyu Han, Wei Jin, Haiyang Zhang, Hui Liu, & Jiliang Tang. 2023. Label-free Node Classification on Graphs with Large Language Models (LLMS).

[11] Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. [Dialogue-Based Relation Extraction](https://aclanthology.org/2020.acl-main.444). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics.

[12] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. [MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations](https://aclanthology.org/P19-1050). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

[1] Bongseok Lee and Yong Suk Choi. 2021. [Graph Based Network with Contextualized Representations of Turns in Dialogue](https://aclanthology.org/2021.emnlp-main.36). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 443–455, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[2] Xiao Liu and Jian Zhang and Heng Zhang and Fuzhao Xue and Yang You. 2023. [Hierarchical Dialogue Understanding with Special Tokens and Turn-level Attention](https://arxiv.org/abs/2305.00262).

[3] Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. [ConvoKit: A Toolkit for the Analysis of Conversations](https://aclanthology.org/2020.sigdial-1.8). In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 57–60, 1st virtual meeting. Association for Computational Linguistics.

[4] Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. [COSMIC: COmmonSense knowledge for eMotion Identification in Conversations](https://aclanthology.org/2020.findings-emnlp.224). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2470–2481, Online. Association for Computational Linguistics.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](https://arxiv.org/abs/1810.04805).

[6] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, & Veselin Stoyanov. (2019). [RoBERTa: A Robustly Optimized BERT Pretraining Approach](https://arxiv.org/abs/1907.11692).

[7] <https://github.com/XiaoxinHe/Awesome-Graph-LLM>

[8] Po-Wei Lin, Shang-Yu Su, and Yun-Nung Chen. 2022. [TREND: Trigger-Enhanced Relation-Extraction Network for Dialogues](https://aclanthology.org/2022.sigdial-1.58). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 623–629, Edinburgh, UK. Association for Computational Linguistics.

[9] Weishi Wang, Steven C.H. Hoi, and Shafiq Joty. 2020. [Response Selection for Multi-Party Conversations with Dynamic Topic Tracking](https://aclanthology.org/2020.emnlp-main.533). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6581–6591, Online. Association for Computational Linguistics.

[10] Zhikai Chen, Haitao Mao, Hongzhi Wen, Haoyu Han, Wei Jin, Haiyang Zhang, Hui Liu, & Jiliang Tang. 2023. Label-free Node Classification on Graphs with Large Language Models (LLMS).

[11] Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. [Dialogue-Based Relation Extraction](https://aclanthology.org/2020.acl-main.444). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics.

[12] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. [MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations](https://aclanthology.org/P19-1050). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 527–536, Florence, Italy. Association for Computational Linguistics.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.