



# North South University

## CSE 499A: SENIOR DESIGN PROJECT I

### ETHICAL AND PROFESSIONAL RESPONSIBILITY

#### PROJECT TITLE

AutoNote: Transformative Meeting Summarization and  
Highlighting Points based on NLP

#### Submitted By:

Name	ID	Section
Md. Saiyem Raiyan	2012468042	10
Zobaer Ahammad Zamil	2021796042	10
Sheikh Mohammed Wali Ullah	2021186042	10
Samia Sultana	2014048042	29

#### Submitted To:

Dr. Mohammad Ashrafuzzaman Khan (AZK)  
Asst. Professor, Department of Electrical & Computer Engineering  
North South University

# Ethical and Professional Responsibility

## Overview

Summarization with Natural Language Processing (NLP) uses NLP techniques to automatically generate concise summaries from extensive textual content such as news articles, scientific papers, and legal documents.



## Details

Ethical proficiency in NLP (Natural Language Processing) summarization refers to developing and applying ethical principles and considerations when designing, training, and deploying automatic text summarization models and systems. It encompasses a range of concerns related to fairness, transparency, bias, privacy, and societal impact in the context of NLP summarization. The Summarization of Ethical Proficiency emphasizes the importance of transparency in NLP summarization, including the algorithm used, data used,

and system limitations. It underlines the need for accuracy and fairness, avoiding bias and omitting important information, respecting privacy, and avoiding deception. The summary also demonstrates responsible use, avoiding deception or misleading others. It also addresses ethical considerations for different NLP summarization projects, such as spreading misinformation in news articles, accurately representing scientific papers, and ensuring completeness in legal documents. Ethical proficiency in Natural Language Processing (NLP) summarization is crucial as it can impact various aspects of society, including journalism, education, and legal documentation. Vital ethical considerations include fairness, transparency, bias mitigation, privacy, context awareness, societal impact, user consent and control, continuous monitoring and iteration, and collaboration with diverse stakeholders. Fairness ensures that summarization systems do not discriminate against specific groups or content, while transparency makes the summarization process more understandable. Bias mitigation prevents the propagation of stereotypes or skewed perspectives. Privacy ensures that sensitive information is not exposed in summaries. Context awareness helps avoid misleading or incomplete summaries. Continuous monitoring and iteration ensure the ethical performance of summarization systems.

## Racial or Harmful problem

Addressing racial and harmful concerns in Natural Language Processing (NLP) is critical to prevent AI systems from repeating prejudices, preconceptions, or discriminatory actions. However, if NLP systems prove racially biased, this threatens their promise to make governments more sensitive and more responsive to citizens' concerns. Especially given the reality of racial inequalities in many states, racial bias in NLP risks deepening existing tensions and perpetuating the feeling shared by many people of color that their government does not represent them. Although language may seem a peripheral part of escalating racial tension, how we use language is fundamental to our ability to relate to each other and participate as equals in a democracy. Any technology that threatens to limit the ability of people of color to express themselves wholly or engage with their government threatens some of their most fundamental civil rights.

One potential problem is that NLP summarization systems might be trained on biased data or harmful stereotypes. It can lead to summaries that are also biased or harmful. For example, a system trained on a dataset of news articles about crime may be more likely to summarize offenses committed by people of color than crimes committed by white people. Another potential problem is that NLP summarization systems might be used to spread misinformation or propaganda. For example, a system could generate fake news articles or summarize scientific papers in a way that misrepresents the findings. Here are some specific considerations for mitigating racial and harmful problems in NLP:

Addressing racial and harmful problems in Natural Language Processing (NLP) is crucial for developing responsible AI systems. Bias detection and mitigation are essential; for example, an AI chatbot may inadvertently generate biased responses about different cultures if not trained on diverse datasets and corrected for biases. Ethical guidelines and review processes are needed to ensure AI isn't used for harmful purposes, such as generating hate speech or misinformation. Fairness metrics help assess outcomes; for instance, a resume screening AI must avoid favoring specific names or genders to ensure fair hiring. Transparency and explainability are crucial, like in AI algorithms for credit scoring or housing, where decisions must be understandable and non-discriminatory. User feedback mechanisms, content moderation, and third-party audits are vital to combat hate speech, as seen in social media platforms. Education and community engagement are essential; organizations like OpenAI involve the public in AI model discussions. Regulatory compliance, such as GDPR for data privacy, is necessary, and continuous improvement is vital to reduce biases and harm in NLP systems.

# Brutality

In the context of NLP (Natural Language Processing) meetings and dialogue summarization, "violence" typically refers to the presence of aggressive, offensive, or inappropriate language or content in the generated summaries. The summarization model has produced an outline that includes content that may be harmful, offensive, or objectionable to readers or participants in the meeting or dialogue. NLP models and technologies can be used in various contexts, including violence-related, such as detecting hate speech, cyberbullying, or harmful content online. It can also be used to analyze and understand text data related to violent events or conflicts. Ethical considerations and guidelines must be followed when using NLP in such contexts to ensure responsible and lawful applications. Violence in NLP meetings and dialogue summarization can manifest in several ways:

**Offensive Language:** The summary may contain improper words, hate speech, or derogatory terms that can harm or upset readers or participants.

**Misrepresentation:** If the summary distorts the original conversation or takes statements out of context, it can lead to misunderstandings or misinterpretations, potentially causing conflicts.

**Bias and Stereotyping:** Summaries that perpetuate biases or stereotypes can be considered a form of violence, as they may reinforce harmful stereotypes or prejudices.

**Inflammatory Content:** Summaries that emphasize confrontational or contentious aspects of a conversation without providing context or nuance can contribute to a hostile atmosphere.

**Privacy Violations:** If the summary exposes sensitive or private information shared during the meeting or dialogue, it can be considered a form of violence against participants' privacy.

The violence shouldn't contain this, which attacks one personally or says so about fake news, threads, or misleads that violate the rules and policy. Addressing violence in NLP meetings and dialogue summarization is crucial for creating safe, respectful, and inclusive communication environments. It involves implementing ethical guidelines and content moderation mechanisms to filter out or appropriately handle violent or inappropriate content in generated summaries. It may require pre-processing the input data, designing the summarization model to avoid generating harmful content, and providing mechanisms for users to report and address problematic summaries. Additionally, involving human reviewers or moderators in the summarization process can help ensure that summaries meet ethical and safety standards.

# Slang Removal

In Natural Language Processing (NLP) meetings and dialogue summarization, dealing with slang words can be challenging, as slang terms are often informal, context-dependent, and vary widely across different regions and communities. If a slang word is more in a text, then there are high chances to take the position of that particular slang word in summary because it works as an attention mask. So it has to be removed. For example, We put a dataset of slang words so that it can be filtered slang or bad-meaning words. Here are some considerations for handling slang words in NLP meetings and dialogue summarization:

**Identifying Slang Terms:** To deal with slang efficiently, NLP models must be educated or equipped to detect slang words and phrases. Building specialized dictionaries or lexicons of slang terminology particular to the domain or location of interest may be required. Machine learning models may also be fine-tuned using slang datasets.

**Normalization and Mapping:** Slang terminologies are frequently linked to their conventional or more formal equivalents. For example, "gonna" can be mapped to "going to" and "wassup" to "what's up." This type of normalization can assist in increasing the accuracy of summarization.

**Contextual Understanding:** NLP models used for dialogue summarization should be able to grasp the context in which slang phrases are employed. Slang terms can have several meanings depending on the context.

**User Preferences and Guidelines:** Slang terms may be kept in summaries in some circumstances to reflect the informality and tone of the original discussion. Others may prefer that slang terms be substituted with more official terminology.

**Evaluation and Fine-Tuning:** It's essential to continually evaluate and fine-tune NLP summarization models to ensure they handle slang words effectively. Feedback from users can be invaluable in this process to improve the quality of summaries.

**Domain-Specific Slang:** There may be domain-specific slang or jargon used in meetings and dialogues depending on the setting (e.g., legal, medical, academic). Recognizing and dealing with such lingo is critical for proper summarization within such fields.

Therefore, use slang words in NLP meetings and dialogue summarization responsibly: Use appropriate slang words in NLP meetings and dialogue summarization, avoid offensive or discriminatory words, define slang words, and use them sparingly when they add value. It's an area that can benefit from ongoing development and refinement to improve the quality of summaries generated by NLP systems.

## Zero Tolerance for Deception

In NLP (Natural Language Processing) meetings and dialogue summarization, the "honest output" concept is fundamental. It denotes the importance of generating summaries containing accurate information and not introducing false or misleading content. Ensuring honest output is essential for maintaining the integrity of communication and decision-making processes in various contexts, from business meetings to legal proceedings, by providing reliable and trustworthy summaries of the discussions and dialogues. This principle underscores the ethical responsibility of NLP systems to prioritize accuracy and truthfulness in their summarization outputs. Ensuring that the NLP meeting and conversation summarization output is honest and accurate is essential. False information can have a negative impact on individuals and society, so it is vital to take steps to mitigate this risk.

In summarizing, the result or output should be relevant according to the main text, dialogue, meeting, etc. The main thing is that the production generates honest information using suitable models, which must be true. False data or information should be discarded. Wrong information meetings and dialogue summarization can negatively impact individuals, organizations, and society. It can spread misinformation about terrorist attacks, scientific findings, or current events, leading to uninformed decisions and public unrest. It's crucial to be aware of this risk. However, to maximize honesty and accuracy, it needed to ensure the accuracy of a Natural Language Processing (NLP) model, use diverse and high-quality training data, implement fact-checking mechanisms, mitigate biases, ensure transparency and explainability, have human oversight, gather user feedback, provide legal and ethical compliance, and continuously evaluate the model's performance. Educate users about the capabilities and limitations of NLP systems, encourage critical thinking, and use NLP technology responsibly, considering its potential impact on society and addressing ethical concerns. Regular updates and fine-tuning can enhance accuracy over time.

While these techniques can improve the honesty and accuracy of NLP outputs, it's crucial to recognize that perfection may not be possible owing to NLP technology's inherent constraints and limits. A mix of these tactics, continuing research, and appropriate use is required to continually increase the dependability of NLP systems in meetings and conversation summaries.

## Discrimination with Fairness

The central theme for the principle "No discrimination, all the same priority" means that summarization models should be trained on and used in a way that does not discriminate against any individual or group of people. There are several ways in which NLP project summarization can lead to discrimination. For example, if a summarization model is trained on a dataset that is biased against certain groups of people, the model is likely to generate summaries that are also biased. Additionally, suppose a summarization model is used to generate summaries of news articles or

other types of text that contain discriminatory language. In that case, the model will likely amplify that language and make it more visible.

Picture a spoken dialog system that is easy for a young male financial professional user with a London English pronunciation but that may barely work for an elderly lady from Uddingston (near Glasgow, Scotland). As automated information systems are becoming more pervasive, they may eventually substitute human information kiosks for cost reasons, and then out-of-sample user groups could be excluded and left behind without an alternative. The internal functioning of NLP systems can raise questions of transparency and accountability: what if a parser does not work for particular types of inputs, and the developer does not communicate this aspect to an application developer who wants to build a medical application that uses it? It is responsible behavior to disclose the limitations of a system to its users, and NLP systems are no exception.

So, it is crucial to be mindful to avoid any kind of discrimination. This principle underscores the importance of treating all individuals and data sources equally and without bias throughout the project's lifecycle.

- **Equal Treatment of Participants:** Autonote and NLP Summarization systems prioritize equal treatment of all meeting participants, irrespective of their identity, background, role, or influence. This approach promotes inclusivity and respect by ensuring that every voice is considered equally important during the summarization process. To maintain fairness, it's crucial not to discriminate against any data source based on linguistic or cultural characteristics. All data sources should receive equal consideration and respect. Using summarization models on dialogue containing discriminatory language, such as racist or sexist jokes, may result in the models amplifying and highlighting such language in the summaries.
- **Objective Content Selection:** Content selection for summaries is based on fair and unbiased criteria to prevent favoritism or distortion of ideas. This approach encourages diverse perspectives and avoids discrimination in data curation for training models. Eliminating biased data is crucial to prevent models from perpetuating stereotypes or unfair behavior. When summarization models handle sensitive topics like race or gender, they might generate one-sided or misleading summaries, particularly in controversial discussions like affirmative action.
- **Impartial Handling of Decisions:** Decisions taken during meetings are treated with unwavering impartiality. These systems focus exclusively on the substance and context of decisions without regard for their origin, nature, or the individuals involved. It ensures that decision summaries are accurate and devoid of bias, fostering transparency and accountability.

- **Mitigating Bias in Model Outputs:** NLP models sometimes produce biased or discriminatory outputs. Implement bias detection and mitigation techniques to ensure the model's responses or summaries do not discriminate against any individual or group.
- **Inclusive Language:** Pay attention to the language used when generating or summarizing text. Avoid using offensive or discriminatory language, and make efforts to be inclusive in the choice of words and phrases.

To avoid discrimination or any biased activity, we should be mindful of the following:

- **Use high-quality data:** When training our summarization models, we should use high-quality datasets representative of the population the models will be used to summarize dialogue. It means avoiding datasets that are biased against certain groups of people.
- **Evaluate the models for fairness:** Once our summarization models have been trained, we should evaluate them for righteousness. This can be done by generating summaries of dialogue that is known to contain discriminatory language and seeing how the models handle that language.
- **Use the models responsibly:** When using our summarization models, we should be mindful of the potential for discrimination. It means avoiding using the models to generate summaries of dialogue that are likely to contain discriminatory language.

## Respecting Privacy: Obtaining Permission for Meeting Transcripts

In the context of our NLP project “Autonote” and the broader field of NLP Summarization, the ethical and professional responsibility of obtaining permission to take meeting transcripts underscores the commitment to respect privacy rights and ensure the responsible use of data.

It is important to obtain permission from all meeting participants before taking the meeting transcripts and using them to train and evaluate our NLP meetings summarization (Autonote) and NLP Summarization. Meeting transcripts can contain sensitive and confidential information, such as trade secrets, personal data, and financial information. Before initiating the collection of meeting transcripts, explicit informed consent is sought from all participants. This consent outlines the purpose of data collection, the use of the transcripts for summarization, and any potential implications. Participants have the right to grant or withhold their consent based on clearly understanding how their data will be utilized.

Transparency is paramount in the process of obtaining permission for data collection. Participants are informed about how their data will be used, who will have access to it, and the safeguards to protect their privacy. Additionally, there is accountability for adhering to privacy regulations and policies. There are a few different ways to obtain permission from meeting participants. One way is to include a statement in the meeting invitation that the meeting will be recorded and that the

transcripts will be used to train and evaluate NLP meetings summarization (Autonote). If participants disagree with the recording, they can decline or leave the meeting invitation.

Another way to obtain permission from meeting participants is to ask for their consent verbally at the beginning of the meeting. This can be done by saying, "This meeting is being recorded for training and evaluating our NLP meetings summarization (Autonote). Please let me know if you do not consent to using your transcript."

Once we have obtained permission from meeting participants, we should protect their privacy. This includes storing the meeting transcripts in a secure location and restricting access to the transcripts to authorized personnel. We should also have a process for deleting meeting transcripts when they are no longer needed.

In addition to the above, we should be transparent about how we use the meeting transcripts. This means letting meeting participants know how their transcripts will be used, and their privacy will be protected. We should also give meeting participants control over their transcripts. For example, meeting participants should be able to request that their transcripts be deleted.

By following these guidelines, we can ensure that NLP meetings summarization (Autonote) is ethical and responsible and respects meeting participants' privacy. In an era where data privacy and security are vital concerns, this strategy develops trust, protects individual privacy, and promotes responsible data handling.

## Explanations with Integrity

In the context of our Natural language processing summarization project, "Autonote," and the broader field of NLP Summarization, the ethical and professional responsibility of maintaining confidentiality and providing explanations that are neither problematic nor illogical is a fundamental principle that underscores our commitment to responsible data handling and clear communication. We must ensure that the meeting transcripts we collect and use are kept confidential and are not shared with anyone who is not authorized to have access to them. Meeting transcripts can contain sensitive and confidential information, such as trade secrets, personal data, and financial information.

We should take the following precautions to preserve confidentiality:

- Keep the meeting transcripts in a safe place. It might be a cloud-based storage service with high encryption or a restricted on-premises server.
- Only authorized people should have access to meeting transcripts. Members of the NLP project team, managers, and other relevant stakeholders may be included.

- Create and put in place a confidentiality policy. This policy should define the methods for managing and keeping meeting transcripts and the repercussions of policy violations.
- Employees should be educated on confidentiality and the protocols for managing meeting transcripts.

## Conclusion

In a broader context, the discussions at the Ethical and Professional Responsibility of NLP meeting epitomize a pivotal moment in the intersection of technology, ethics, and society. The convergence of Natural Language Processing (NLP) with real-world applications has brought forth a host of ethical dilemmas and professional responsibilities that demand our immediate attention. A central theme of the meeting was the importance of moral awareness. Stakeholders recognized that NLP practitioners must be acutely aware of the ethical implications of their work. It involves addressing bias, privacy concerns, and spreading misinformation. The consensus was that ethical mindfulness should be integral to NLP research and development. These challenges encompass biases ingrained in algorithms, potential privacy infringements, and the dissemination of misinformation. The meeting consensus underscored the need for a robust framework that outlines the ethical boundaries of NLP development and deployment. Such guidelines should encompass a range of issues, including bias mitigation, transparency in algorithmic decision-making, and responsible data management.

Furthermore, the meeting emphasized the importance of accountability. As NLP systems become integral parts of various industries, there is a growing imperative to hold individuals and organizations accountable for their actions. This includes acknowledging the consequences of biased algorithms and ensuring that ethical considerations are integral to NLP research and development.

The meeting emphasized the importance of transparency, diversity, and collaboration in the ethical and professional responsibility of Natural Language Processing (NLP). It demonstrates the need for NLP systems to provide clear explanations for decisions, fostering user trust and understanding. It also stressed the need for collaboration among academia, industry, and policymakers to establish ethical standards and adapt to rapidly evolving technologies.

In conclusion, the "Ethical and Professional Responsibility of NLP" meeting revolved around the imperative to address ethical concerns in the field. Key takeaways encompass ethical awareness, guidelines, accountability, transparency, diversity, collaboration, and education. The overarching message was that ethical considerations must be central to NLP advancements to ensure these technologies contribute positively to society while minimizing potential harm.

## References:

- I. [Shared\\_racial\\_bias\\_in\\_natural\\_language\\_processing\\_-\\_report.pdf](#)
- II. [Understanding Ethics in NLP Authoring and Reviewing](#) Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744.  
<https://doi.org/10.1007/s11042-022-13428-4>
- III. Yang, A. (2023, April 9). Critical summary on the ethical use of NLP - Amy Yang - Medium. *Medium*.
- IV. [Critical Summary on the Ethical Use of NLP | by Amy Yang | Medium](#)
- V. <https://aclanthology.org/W17-1604.pdf>