Speech Act Theory and Ethics of Speech Processing as Distinct Stages: the ethics of collecting, contextualizing and the releasing of (speech) data

Jolly Thomas, Lalaram Arya, Mubarak Hussain, S. R. Mahadeva Prasanna
Indian Institute of Technology Dharwad, Dharwad-580011, India
Email: joliethomas@iitdh.ac.in, 202021004@iitdh.ac.in, 193101002@iitdh.ac.in, prasanna@iitdh.ac.in

2023 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS) | 978-1-6654-5713-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ETHICS57328.2023.10154932

***Abstract*—Using speech act theory from the Philosophy of Language, this paper attempts to develop an ethical framework for the phenomenon of speech processing. We use the concepts of the illocutionary force and the illocutionary content of a speech act to explain the ethics of speech processing. By emphasizing the different stages involved in speech processing, we explore the distinct ethical issues that arise in relation to each stage. Input, processing, and output are the different ethically relevant stages under which a spoken item or a speech navigates within the range of speech-processing modules. Employing the illocutionary force-content distinction, we specify and characterize the input-related ethical issues, the output-related ethical issues, and the processing-related ethical issues involved in speech processing. Together with illocutionary force-content distinction, we employ the data-information distinction to characterize the stage-wise ethical issues in the phenomenon of speech processing as the ethics of collecting (speech) data, the ethics of contextualizing (speech) data/information, and the ethics of releasing the contextualized information (processed speech). Immediate ethical issues that arise from the range of speech processing modules are distinguished from distant ethical issues. We also indicate the nature of ethical issues that arise from Speaker Independent speech technologies.**

***Index Terms*—Speech acts, Ethics of speech data, AI ethics, Ethics of speech processing, Privacy, Ethics of illocutionary force**

I. Introduction

he rapid development in speech technology gives rise to novel ethical issues. The development of any technology relating to speech is dependent on the very domain of speech processing. In this paper, we introduce certain concepts that are relevant for developing an ethical framework for speech processing systems. Rather than providing any ethical guidelines, this paper aims to provide a framework to think about the nature of the ethical issues arising in the domain of speech processing. An ethical issue whether relating to a deep-fake technology or to a speech synthesizer or a speech recognition technology is better addressed within the large speech processing framework of which these miniatures are part. The functioning of any speech technology is primarily dependent on some form of speech processing. Whether it is an Automatic Speech Recognition (ASR) or an Intelligent Virtual Assistant (IVA) or any technology relating to Speech Emotion Recognition in Conversation (ERC), all of them primarily function in the speech processing system. Thus, we focus on the ethics of the entire range of speech-processing modules and not on any particular technology that is emergent in this process. To explore the nature of ethical issues, we emphasize the very **process** aspect of the speech processing domain as having different stages in this process.

II. Different stages of the speech processing for ethical/moral analysis

Under the entire range of speech processing modules, a spoken item or a speech could go through at least three stages in a manner relevant to a moral agent giving rise to certain moral issues. Those stages are input, processing, and output. The input stage within the range of speech processing modules concerns the collection of the data which are in the form of spoken items of a given speaker. In some cases, we may consider speech recognition as the aspect that is relevant at the input stage in speech processing. However, speech recognition itself cannot be an exclusive aspect of the input stage as speech recognition techniques might often fall in the output stage too. For studying the nature of ethical matters, let us count something that happens prior to speech processing as the first stage: the collection or acquisition of speech. The second stage, which is the processing stage involves the manipulation of the collected speech through some synthesizing method that brings forth the output of a certain sort to meet a purpose. Apart from the spoken item, the text could also be the input for the processing stage. In the third stage, what we have is the output because of some speech processing system: a released speech or a spoken item, or a text that reaches at least a moral agent whose actions are impacted. What is central to the input stage–a spoken item or a text, and the output stage–a spoken item or text that makes an impact on at least a moral agent, is the very activity of the processing of the speech. Taking the entire range of speech processing modules as a whole into account, we develop an ethical framework for speech processing.

The emphasis of the stage-wise division of the range of speech processing modules provides a possibility to look at the nature of different kinds of ethical issues that are relevant to each stage. In the speech processing phenomenon, some of the ethical issues that are relevant to a specific stage **S** might be distinct in the sense that even without considering how **S** is related to other stages such as **S1** and **S2**, the **S** might give rise to the ethical issues of its own kind. The entire range of

978-1-6654-5713-2/23/$31.00 ©2023 IEEE

speech processing modules might give rise to some relevant ethical issues. The entire range of speech processing modules in their combination of different stages that solely give rise to ethical issues. Also, each stage gives rise to some relevant ethical issues as ethics concerning that stage. Thus, we have distinct ethical issues pertaining to each stage in the following manner: ethical issues that arise due to (a) the inputting of speech or text of some sort, (b) the processing of the collected speech, and (c) the outputting of speech or text to meet certain purposes. In the following section, let us see how the different stage-wise ethical issues arise due to the kind of relations held between the modules of speech processing and the moral agent.

1) **Moral agent, input-output output-input relations, and differences in the ethical issues**
In the case of speech processing, how different modules in speech processing and a moral agent are related impacts the kind of ethical issues that arise. At the stage of input, there is some input-confined relation between the speech of a moral agent **A** and a system **S** or a particular module that collects the speech of moral agent **A**. Also, at the stage of output, there is some output-confined relation between the speech of a system **S** from which a speech (bearing some similarity with a moral agent) is released and a moral agent **A** to whom the released speech makes an effect or an impact. In addition, there are ethical issues concerning the very activity of speech processing itself, and speech processing holds an intermediary relation between input-confined relation and output-confined relation. The point that we emphasize in this paper even if a trivial one is the following: the way the input-confined relation impacts a moral agent is different from the way the output-confined relation impacts a moral agent. Therefore, we need to distinguish between the ethical issues arising out of input-confined relations, and processing and output-confined relations as different kinds.
The input-confined ethical issues points towards both collection of data and its purpose. Any system used to collect the data for the speech/voice recognition system is relevant for the input-confined ethical issues, as they primarily are concerned with the collection of speech data to meet certain purposes. Inputting is mostly given a teleological explanation: certain speech data is collected for some purpose or goal. To some extent, the teleological explanation might even give moral justification for the speech recognition activity itself. Output-confined ethical issues arise once the processed speech is out making an effect on at least a moral agent. Whatever we listen to from Alexa and whatever changes we make in the course of our actions is an example of this. If the actions turn out to be ethically significant then there arise ethical issues directed upon the activity of releasing the processed speech. Outputting of the processed speech too will have a teleological explanation as in the case of input-confined ethical issues. The very release of the processed speech is for some purpose and sometimes the purpose itself might give moral justification for releasing the processed speech through some processing systems or synthesizers. The entire range of speech-processing phenomena assumes a teleological path: speech is collected, processed, and released for some purpose. In this teleological path, sometimes humans are the subjects whose purpose is to be served, and many times humans are objects of the purpose.
2) **Input-confined and output-confined ethical issues: Deepfake Technologies** [14]
Let us consider speech-related deepfake technologies as an example to understand the distinction between input-confined ethical issues and output-confined ethical issues. What the deepfake technology does to a human is intrinsically wrong or instrumentally wrong. There are studies indicating the ethical issues that arise from deepfake technologies [6]. Ethical issues arising from the deepfake technologies that fall at both the input stage and output stage can be given the following kind of explanation. There is nothing intrinsically or inherently wrong in using or in the applications of deepfake technologies. Addressing the input-confined ethical issue or eliminating the morally wrong actions at the stage of input depends on the satisfaction of the consent requirement. The only way to satisfy the consent requirement is to ensure that the person from whom the personal speech data is collected with the consent to use the data, must be made aware of the entire teleological path of the purpose that is realized through the deepfake technology. Making aware of the entire teleological path is a necessary condition for the satisfaction of the consent requirement. To address the input-confined ethical issue, the user needs to be made aware of the intent of the application of a particular deepfake technology in the given scenario. Addressing the input and the output confined ethical issue presupposes the processing of the speech data in a specific manner. At the stage of processing, one may not tamper with the expected type of processing with the strong consideration of the avoidance of input and the output confined ethical issues. It means the processing of speech data is absolutely restricted or constrained by the input and the output-confined ethical issues. Thus, the processing of speech data is morally acceptable only to the extent that the processing avoids both input and output-confined ethical issues.
3) **Self-learning aspect of AI systems: DeepMind, Medical AI, and Artificial General Intelligence (AGI)**
Consider the self-learning aspect of AI systems or models. There are different cases that indicate AI systems can develop the capacity to learn by themselves. A particular AI model might be trained to perform a specific task, subsequently, it could so happen that the AI model could start to perform some novel and unintended or unexpected tasks. Such a self-learning aspect promotes

stronger autonomy for the AI model. Alpha Go, which was developed by DeepMind can be an example here. AlphaGo which is based on Deep Neural Networks is trained through reinforcement learning. Because of this, the AI algorithm can improve itself over a period. In 2015 AlphaGo defeated the world go champion Fan Hu. Under what ethics could the above-mentioned cases fail? In 2017, DeepMind developed AlphaGo into AlphaGo Zero. The self-learning aspect is even stronger in AlphaGo Zero. It can learn even certain rules of the game by itself [1] [16]. This is a case where a particular AI model goes beyond the intended task and starts performing novel functions.

The self-learning aspect is obvious in the case of AlphaGo. Like the AlphaGo case, a problem relevant to the scenario of speech processing can be considered from Medical AI. Speech is extensively used in the case of healthcare applications. For example, in the case of Parkinson's, speech signals are used to detect and monitor the disease. It was observed that such AI models, even if meant to monitor the disease, could gather, or learn some additional information regarding gender, speech tasks, native language, etc [13]. Self-learning aspects of additional information in which the systems themselves enjoy stronger autonomy is indeed a relevant ethical issue. The use of Health Recommender Systems (HRS) which again employs speech processing techniques affects the autonomy of patients [19]. Another example could be cited from the discussions regarding Artificial General Intelligence (AGI) which is much different from narrow AI. The level of autonomy involved in the case of AGI is that the intervention of humans is nil [18]. Such cases can be considered even in the phenomenon of speech processing.

Keeping the current framework, the following questions can be asked. Under what ethics the above-mentioned cases could fall? Should one consider them input-confined ethical issues or as output confined ethical issues? Does the ethical problem fall in the processing stage?

4) **Intelligent Personal Assistants (IPA)**

The ethical concerns arising from the use of Intelligent Personal Assistants (IPA) can also be brought under the purview of the ethical framework developed in this paper. Apart from privacy and consent related issues, data ownership becomes a crucial ethical concern, especially in the case of IPAs [9]. The court case in Arkansas could be considered here [11]. Prosecutors are demanding to share the recordings from Amazon Echo for the murder investigation. In such scenarios, both questions of ownership and trust come together. Who owns the agent's data? As a consumer, how would I be able to trust the IPA technology that could testify against me? The way the speech data would be used (a) must be made aware to the one from whom the speech data is collected and (b) the purpose for which the data is collected should also be made aware. The court case in Arkansas clearly shows that the entire teleological path cannot be conceived and made available to the consumer. Both input-confined and output-confined ethical issues arise in the case of IPA.

III. On the nature of speech: Speech act theory, Illocutionary Force, and Illocutionary Content

Let us consider the explanation given to the notion of speech or utterance or speech act or illocutionary act [4] [7]. Minimally, one could say that the act of speech is nothing but the act of utterance by a speaker. Such utterances are expected to make some sense to a listener. Speech acts are those acts that one could perform by saying or uttering that she is doing so. For the utterance "I promise" is a speech act in the sense I am performing a particular act of "promise making" by saying that "I promise". In a speech act, both the utterance of a sentence and the performance of the action that is mentioned in the sentence come together. The speech act of a speaker not only provides some information from the speaker but also action performed by the speaker. Another concept discussed in the speech act theories concerns the performative sentences: they are the "first person, present tense, indicative mood, active voice, that describes its speaker as performing a speech act" [7]. I assert that Tom cheated on the exam. Here the speaker performs the action of asserting something by saying that she is doing so (performing the act of assertion). It is possible for the speaker to perform a speech act without expressing a performative sentence. The speaker can assert that Tom cheated on the exam without saying that she **asserts** that so and so is the case. It is also possible for the speaker to utter the performative without the performance or without intending the performance of the speech act. For example, I can make a promise using performative without intending to keep the promise or without performing the speech act. There are many speech acts that do not use the performatives: the speaker's speech acts without the utterance of performatives. There are many performatives without intending the speech act: speaker's performatives without intending to perform the speech act. Illocution is another concept that is used synonymously with speech acts. Illocution is the act of speaking something such that the very speaking itself is constitutive of the action that is intended by the speaker. The very utterance constitutes the action itself. Speech acts are illocutionary acts. As we have illocutionary acts, we have locutionary acts. Locutionary acts are the utterances of the sentence in its literal sense. For example, if someone asks, "Do you have a minute?", the illocutionary act is "Please, I would like to talk to you." However, the locutionary act is a question and its literal sense is irrelevant to the speaker, and also for the listener.

A very significant distinction to be considered here is the distinction between illocutionary force and illocutionary content. Illocutionary content is what the speaker says. The content of what the speaker says must be distinguished from the force that the speaker has on what she says. The literal

meaning is not the only constituent of the utterance or speech. Force of the speech or the force of the utterance is also an aspect of the illocutionary act or speech act. We use illocutionary acts and speech acts synonymously. The mere content of the speech act is the proposition or what the sentence expresses literally. The illocutionary force which is an aspect of the speech or illocutionary act is the achievement aspect that the speaker has on the listener by having a speech illocutionary act. Searl and Vanderveken [15] bring clarity into the notion of illocutionary force providing the following features of illocutionary force.

1) Illocutionary point: it is the objective of a particular speech act. Each speech act has a specific objective corresponding to that speech act. The objective of the speech act of describing is to describe and similarly the objective of the speech act of promising is to make a promise.
2) Degree of the strength of the illocutionary point: it is possible for different illocutionary acts to have the same illocutionary point but different strengths. This variation is due to the variation in the illocutionary force. The illocutionary force involved in making a request is different from the illocutionary force involved in making a command. For example, the strength of the illocutionary point is different from saying that "I request you not to visit my office often." from saying "I implore you not to visit my office often."
3) Mode of achievement: the way an illocutionary point is realized can vary. I am being the owner of my house; I can assert by saying that "I assert that you are the caretaker for my house." At the same time, by being the owner of my house I can also appoint by saying that "I appoint you as the caretaker of my house." Here, the illocutionary point is the same but the way the illocutionary point is realized is different in the above-mentioned illocutionary acts.
4) Content conditions: some illocutions have their own propositional content. Therefore, the mode in which some of the illocutions can be achieved depends on the propositional content. The illocutionary point of my promise-making can only be achieved if the propositional content is appropriate, i.e., if the propositional content is about only those that are under my control.
5) Preparatory conditions: all the relevant conditions required for the success of a speech act.
6) Sincerity conditions: this condition has something to do with the psychological state that a speaker should have at the time of the speech act needs. When I make a promise, I need to have the psychological state of the promise being kept or the intention to keep the promise. Or else, the sincerity condition is not satisfied. Similarly, each illocutionary act has a respective sincerity condition corresponding to a psychological state.
7) Degree of the strength of sincerity conditions: the strength of the sincerity condition might vary depending on the speech or illocutionary act.

In this context, we can consider what Grice calls non-natural meaning or the speaker meaning [8]. Not all the aspects of my speech act are connected to the content of my speech act. For example, the volume of my speech has nothing to do with the content or the meaning of the content of what I speak. However, the force is indeed an aspect that constitutes the force. Grice calls the relevant aspect that contributes to the meaning of speech act as non-natural meaning or speaker meaning.

Grice holds that for speaker meaning to occur, not only must one (a) intend to produce an effect on an audience, and (b) intend that this very intention be recognized by that audience, but also (c) intend this effect on the audience to be produced at least in part by their recognition of the speaker's intention. According to Grice, the speaker meaning happens at least in the following three manners.

1) Intention to produce an x impact on the listener
2) Intention that the intention to produce an x impact on the listener be recognized
3) Intention that the x impact on the listener is to be produced by the listener's recognition of the speaker's intention

IV. Application of Speech Act theory to explain the ethics of speech processing

We briefly looked at the theory of illocutionary force and the concept of occurrence of speaker meaning and the notion of the speaker's intention which will give a basis to explicate the ethical issues involved in speech processing. If the above mentioned are the features of an actual utterance made by a speaker or a speech act or illocutionary act by a speaker, then certain ethical issues do arise with regard to the use of speech data that is collected, processed, and released. To explain this, we can mainly consider different features of the illocutionary force. When an actual speech/illocutionary act is performed, at least some of the above-mentioned features of illocutionary force and also some of the features of the speaker's intention are satisfied by that actual speech act. Some features might not be relevant in some contexts. Any speech act is embedded in a particular context or is context-bound. There is no context-free speech act by a speaker. A speech act is context-dependent and the features of illocutionary force are also determined by the context of the utterance. We reiterate the trivial point: any features of illocutionary force of a speech act that is contextual wholly belong to the speaker(s) of that context. It is possible for the illocutionary content of a speech act to remain the same across different contexts of utterance. However, it is highly improbable for the illocutionary force of a speaker's illocutionary act to remain the same across contexts. Across different contexts, even the same speaker cannot execute a speech act with the same illocutionary force but can perform a speech act with the same illocutionary content. Within a context, the speaker's psychological state and the external conditions that the speaker is in determine the illocutionary force of that speaker's speech act. Though the illocutionary

content could remain the same across various contexts, the illocutionary force cannot remain the same across contexts. And the same goes for Grice's intention that the speaker has: the context of the utterance determines the type of intention that the speaker has at the time of the speech act. Whether the illocutionary content of a speech act is context dependent or not might be a contentious matter. But both illocutionary force and intentions are context dependent.

In the scenario of speech processing, what grounds the ethical concerns? Let us consider the example of privacy-related issues. Regarding the collection of speech data, the significant question would be the following. When a set of speech data is collected from a speaker, what is being collected: the illocutionary force of a speech act or the illocutionary content of a speech act? We emphasize that when the speech data is collected, it is not only the illocutionary content but also the illocutionary force of the speech act that is collected. Therefore, in the case of speech data, at least there is a concern to detach the illocutionary content from the illocutionary force. More than the illocutionary content, it is the collection of the illocutionary force of the speech act that intensifies the ethical concerns. Because there is always indeterminacy in detaching the content from the force. What we regard as speech data contains not merely the illocutionary content of the speech act. Together with illocutionary content, the speech data always encompasses the illocutionary force which essentially is tied to the speaker at the time of the speech act. As we have seen previously the illocutionary content might remain the same however the associated illocutionary force can never remain the same across contexts even with the same speaker. More than the illocutionary content, it is such illocutionary force that is relevant to be the speech data.

To make the analysis of the ethics of speech data more precise, the following distinction needs to be considered: the ethics concerning the illocutionary force and the ethics concerning the illocutionary content. Let us see how this distinction explains the ethical issue that is unique to the speech data. Regarding the collection of speech data, the explicit ethical issue that is often pointed out is privacy-related issues. When, as a speaker, my speech data is collected then it fundamentally affects my privacy-related concerns. Privacy issues are not something unique to the collection of speech data. Any personal kind of data will give rise to some privacy-related ethical issues and not just the speech data of a speaker. How to explain the distinctness of the ethical issues that results from the collection of speech data (which is personal kind) as opposed to the ethical issues that result from the collection of other personal kind data? In the scenario of speech processing, in what sense does the collection of speech data affect privacy concerns? First, we emphasize that any ethical issue in the case of speech data arises not just from illocutionary content. The uniqueness of the ethical issues in the case of speech data collection resides in indicating that mostly the ethical issues result from the illocutionary force than from the illocutionary content. To explain this difference of force-content distinction in the case of privacy-related issues, let us consider the suitable features of the illocutionary force of a speech act that we had discussed previously. Consider the 6th and 7th features. A speaker's speech act takes place with sincerity condition: a (psychological) state of the attitude and the degree of the strength of the attitude of sincerity. Each speech act co-occurs with a state of the attitude of sincerity with varying strengths of the attitude of sincerity. In a speech act, to a speaker, there is a strong sense of belongingness of such a state of attitude. Now when that speech act is collected a particular state of the attitude of sincerity which essentially belongs to that speaker is completely severed from that speaker and the context of the speech act. The speaker, the sole possessor of the state of the attitude of the speech act is completely ripped of its ownership. About this specific case, one could say that the severity of the ethical issue seems to be stronger in collecting the illocutionary force than the illocutionary content. In the speech act, the issue seems to be more resilient on the very possibility of collecting the state of the attitude of sincerity a feature of illocutionary force than on the possibility of collecting the illocutionary content.

The issue might become more insistent when such speech data are allowed to be processed and released in a different situation or context. Upon the 7th feature of illocutionary force, we can say that the degree of strength of the sincerity of the illocutionary force is solely possessed by the speaker. It is not the content that plays a crucial role in giving rise to the privacy issue and one of the reasons for the privacy-related issues is the degree of the strength of the sincerity being violated. Speech acts with a state of sincerity having some degree of strength that essentially belongs to the speaker, unlike with illocutionary content. Further explanation about the essential belongingness between certain features of the illocutionary force (maybe the features of sincerity or strength of the sincerity or some other features) and speaker is a subject matter that is to be dealt with by the ontology and need not be addressed here. It does not make much sense to say that the illocutionary content belongs to the speaker. Whether the illocutionary content of a speech act belongs to a speaker or not may be a contentious matter but, it is definitely, not a straightforward connection that can be accepted. However, the connection or the relation between a speech act's illocutionary force and the speaker is too strong to deny. Therefore, we say that the illocutionary force of the speech act belongs to the speaker. Unless the speech data collection takes this force-content distinction making the following questions clear, there will always be indeterminacy in fixing the type of data: (a) is it collecting illocutionary content or (b) is it collecting the illocutionary force and (c) what type (force with a feature) of illocutionary force is collected (d) how many types of illocutionary force is collected?

Again, from the 2^{nd} feature of the illocutionary force, we can say that the degree of the strength of the illocutionary point is something that only the speaker can have or possess. The point of possession plays a crucial role in giving rise to privacy-related issues. It is not the illocutionary content that works as a reason for the privacy-related issues, but it

is the removal of the ownership of the degree of the strength of the illocutionary force from the speaker that gives rise to the privacy-related issues. How privacy-related ethical issues arise could be explained by considering the 2^{nd}, 6^{th}, and 7^{th} features of illocutionary force. To this, if we consider Grice's conditions for the speaker meaning then it is clear that each speech act presupposes certain intentions. Once the illocutionary act is collected the intentions which completely belong to the speaker are also collected. Such intentions of the speaker could undergo or are vulnerable to very strong analysis. There is an indeterminacy associated with the collection of speech data as it is difficult to fix what is being collected: the illocutionary content or the illocutionary force. A similar question would arise in the case of speaker meaning: whether it is only the content of the utterance or together with the content the intentions involved in the speech act are collected?

Assuming that taking consent from the speaker will address the ethical concerns here does not really take the complexity of the ethical issues involved in speech processing into account. From the force-content distinction, we say, there is always indeterminacy in indicating for what the consent is sought: is it for illocutionary force or for content? Regarding the processing stage, the question is, what type of processing such an illocutionary force would undergo? A deviation from any features of illocutionary force (the illocutionary point or mode of achievement or preparatory activities) of the speech act or illocutionary act clearly raises some ethical concerns. Speech data are the actual utterances of the speaker. The content of the utterance is definitely important data. However, what makes speech data unique from other kinds of data is that they are the actual speech act or illocutionary act, or actual utterance. The mode in which the utterances are made is definitely the part of utterance or speech act. One of the ways to explicate the significant factors involved in the mode in which the speech act is to depend on the speech act theory, especially, to look at the illocutionary forces involved in the illocutionary act. It is the actual utterances that undergo the processing. For example, to conduct some studies, we can collect data or information using a questionnaire. If we were to collect the same data or information using a voice recorder, it no more remains the same data. The data collected through the voice recorder will be the same data as that of the data collected through the questionnaire and the sameness lies only at the level of illocutionary content. In the case of the voice recorder, the actual speech act of the speaker from whom the speech data is collected, the illocutionary force also becomes the part of data. Illocutionary force gives better access to the intentions of the speaker than illocutionary content. Access to the speaker's intention and character of the speaker is less difficult with illocutionary force than the illocutionary content. Therefore, speech data gives better access to the speaker's intentions and personal traits. Speech data is unique in that sense. Or at least speech data must be distinguished from other kinds of personal data.

V. Data-information distinction and the illocutionary force-content distinction

Data-information [2] distinction is considered to explicate some of the nature of the ethical issues arising in the case of speech processing. Data are the facts that are the raw material that the system (human or otherwise) accumulates through observation. Data themselves are useless unless they become information or at least potential enough to transform into information. Thus, information is the contextualized data suitable to satisfy certain purposes arising relative to a context. (a) The speech data and (b) the corresponding information that is resultant of the contextualization of speech data and (c) the mechanism that carries out the data-information transformation are the focus here. The very activity of turning data into some contextualized information, or in other words, the transformation of data into information presupposes a mechanism, a method, or a system that carries out such transformation. In the case of speech processing, as we know, the mechanism that performs data-information transformation is the technique of speech processing. In the case of speech, the data-information distinction remains a fundamental distinction that characterizes the distinct stage-wise ethical issues resulting from the phenomenon of speech processing. The illocutionary force-content distinction is to be linked with the data-information distinction. Then we have the illocutionary force in relation to some set of speech data and the illocutionary force in relation to contextualized information. Similarly, illocutionary content in relation to some set of speech data and illocutionary content in relation to contextualized information. Accordingly, the ethical issues are to be distinguished. In the scenario of speech processing, the ethical issues that arise from the illocutionary force relating to the data alone is to be separated from the ethical issues that arise from the illocutionary force relating to the information. Similar distinctions are to be maintained in the case of illocutionary content. That is the ethical issues that arise from the illocutionary content relating to the data alone is to be separated from the ethical issues that arise from the illocutionary content relating to the contextualized information. Let us see this in the following section in detail.

VI. Ethics concerning contextualizing (speech) data, releasing of the contextualized data/information, and the illocutionary force-content distinction

At the input stage, as it was mentioned before, the raw speech data that is collected from a speaker becomes a matter of ethical concern due to privacy-related issues. We should take note of the repercussions of the illocutionary force-content distinction prevalent in the privacy-related ethical concerns in the case of speech data. Even if the raw data does not transform into any type of information, the very activity of collection of speech data per se has some privacy-related ethical concerns. This is so because there is always indeterminacy in distinguishing between the item that is collected in the case of speech data: is it illocutionary force or content of speech act? Ethical issues arising from the speech data and the ethical

issues arising from information corresponding to speech data are of different sorts and are to be addressed separately. In some cases, any data whether it is of speech or otherwise becomes relevant only when the data is potential enough to become some contextualized information. Transformation of speech data to some contextualized information is always a possibility that is inherent to the working mechanism of speech processing. If the entire working mechanism of speech processing does not aim at any form of transformation of speech data to some contextualized information, then the phenomenon of speech processing turns out to be pointless. The very pursuit of speech processing is intrinsically associated with the very possibility of speech data becoming some information relevant to a context. Once the phenomenon of speech processing becomes the subject matter for the domain of ethics, the very activity of speech processing is to be reckoned as the activity of contextualizing the information. What speech processing aims for is nothing but the alteration of the nature of speech data. The relevant explanation of the technique of speech processing for the studies relating to ethics is the following: what the processing does at the end is nothing but the contextualization of the collected data to satisfy certain contextual requirements. Therefore, keeping the entire speech processing phenomenon, whilst we speak of the ethics of collection of data at the initial inputting stage, we speak of the ethics of transformation or the contextualization of speech data at the second stage which is at the processing stage.

The ethically relevant question that arises regarding the contextualization of data is the following. What does speech processing aim to contextualize? Is it illocutionary force or illocutionary content? As was mentioned previously, there is always difficulty in determining whether it is the force or the content that is contextualized. Suppose it is the illocutionary force that is transformed into information, then how to provide moral justification for it? For example, consider the 6^{th} and 7^{th} features of the illocutionary force. Suppose the speech data that is collected were to include the state of the attitude of sincerity (mentioned both in the 7^{th} and 7^{th} features of force) in some form and the processing is to be performed on the collected speech data, then the following question arise. How is the contextualization of the speech data be morally justified? To ask in a different way, how is the collection of my certain state of attitude morally justified for the transformation into some information or for the contextualization? Similar questions could be asked about the other features of the illocutionary force. How is the collection of any features of illocutionary force be morally justified for the contextualization of information? We emphasize that it is not merely the illocutionary content that is transformed into some information or undergoes some processing. But it is also the illocutionary force that undergoes some contextualization. Together with content, my state of attitude, my illocutionary point, the strength of the attitude, etc. become relevant data for the information.

It is not merely the collection of data that requires ethical justification but the transformation of data through some speech processing mechanism— the transformation that the speech undergoes will also require ethical justification. An agent's speech data is collected and the same agent's speech data (either the illocutionary force or the content) will undergo some process of transitions to become some contextualized information. We need to emphasize the ethics of some data **becoming** some information: the speech data itself undergoing some transformation of becoming contextualized information. Within that, we should also focus on the transformation of illocutionary force into some information. Thus, we need to consider the ethics of the transition of data to information. It is not that raw data alone has its role in giving rise to ethical problems. Deep ethical issues arise once the data undergoes the transition of becoming information. Why does the transformation of speech data to contextualized information result in a deeply ethical issue? What exactly does the mechanism of speech processing do to speech data? We should address this question in a manner relevant to ethical considerations. Some metaphysics or ontology is required to look at the nature of speech [10].

Fundamentally, the technique of speech processing alters the very nature of the speech data of an agent. The speech of an agent is intrinsically related to the agent herself/himself or related uniquely in the sense that no voice or speech of an agent is identical to the voice or speech of any other agent in any manner. This claim could be understood in two different ways. One way to understand the intrinsic relation is to say that the physical properties of a speech/sound is unique and are not identical to any other speech/sound. The other way to understand the intrinsic relation is to consider the point of illocutionary force. At least some of the features of the illocutionary force of a speech act of a speaker cannot be reproduced even by the same speaker in different contexts which was discussed in section IV. The technique of speech processing alters the nature of the speech data of a speaker by altering the physical properties and by altering the illocutionary force. In the speech processing phenomenon, it is such speech data of the agent (that has essential relation with an agent) that is allowed to be altered or allowed to be transformed into some contextualized information. Such transformation of speech data or contextualized information precipitates various kinds of deep ethical problems. How such alteration or the transition of data to the contextualized information be ethically justified turns out to be a very relevant question. Thus, we have the question of the ethics of transformation of speech data into contextualized information.

Moving away from the processing stage, once the contextualized information is released to a moral agent, making an effect on an agent does give rise to different types of ethical issues at the output stage. Thus, we have the question relating to the ethics of releasing processed speech or contextualized information. Once the contextualized information is released to an agent, then it gives rise to yet other kinds of ethical issues. Such ethical issues depend on the kind of effect the released contextualized information has on the agent. Agents' course of action and behavior is affected when the

released contextualized information reaches the agents. Thus, we have the question of the ethics concerning the releasing of contextualized information.

One important point to be noted here is that the different types of contextualization or processing of data are possible depending on the requirements at different times. The same data could become relevant at different times depending on the contextual requirements. It is not that the collected speech data will have only one type of processing or contextualization. The collected speech is always vulnerable to multiple contextualizations in the sense that the collected speech can become contextualized information for different contexts at different times. Let us name this problem as data's multiple contextualization vulnerability. The very possibility of data getting stored and transferred makes the ethical issue even deeper. The storage and transfer of data have some ethical considerations as the data being stored has the very possibility of becoming contextualized information at any given point in time. Multiple contextualization vulnerability becomes even more intense or severe when the storage and transfer possibilities of the data are considered.

- Stage 1: Ethics of Collecting (Speech) Data
- Stage 2: Ethics of Contextualizing (Speech) Data/information
- Stage 3: Ethics of releasing the contextualized information (Processed Speech)

VII. Foundation in Normative Ethics: Deontology or Utilitarianism?

Which normative ethical framework would be useful to provide a basis for the ethics of speech processing phenomenon? Which normative framework would justify (a) the ethics of collecting (speech) data, (b) the ethics of contextualizing (speech) data/information, and (c) the ethics of releasing contextualized information (processed speech)? Mostly, the utilitarian ethical framework [12] seems to be useful: speech recognition activity or collection of speech data has some purpose and if the purpose maximizes the utility, then speech recognition activity or collection of speech data is ethically justified. Similarly, the ethics concerning each stage could be justified further using utilitarianism. If we consider different versions of utilitarianism, then we may choose to act utilitarianism vs rule utilitarianism. Or the distinction between extreme and modest kind of utilitarianism [17] might be a requirement here. As the speech processing domain is a new phenomenon that a moral agent faces, the adoption of moral rules governing any activity relating to speech processing is indeterminate and unfamiliar to any moral agent. One may not even know the relevant moral rule that would maximize utility whether in input-confined or output-confined ethical issues. Therefore, act utilitarianism will play a role in assessing morality in the speech-processing domain. It is through trial and error; we learn which moral rule is to be adopted in a specific situation focusing on a particular act. An extreme form of utilitarianism or act utilitarianism can accommodate the ethical issues emerging from the new and rapidly changing technologies such as speech technologies, deepfake technologies, any AI technologies, etc. A suitable ethical framework would be one that takes frequently shifting ethical issues emerging from the very novel technological phenomenon into consideration. In most cases, we cannot have the rule and then introduce the technology. For example, we cannot enlist all the rules and regulations associated with the introduction and use of deepfake technologies, then introduce it in society. Rather, what happens here is that technology emerges, and then, society becomes aware of its ethical issues and the way these ethical issues are to be handled. In such cases, we cannot have the enlisted rules and regulations. Once the novel technology begins to be part of social life, there emerge new types of ethical issues and accordingly, there emerge new rules. In such a situation, the normative ethical framework should be able to consider the specific case or act and be able to assess the morality behind such case or act. The rules might change from time to time.

From a deontological ethics point of view, it can be argued that personal speech data collection in itself is morally wrong or intrinsically wrong. It is so because the collection of any personal data is a case of violation of a person's right to privacy. Even if the such collection of personal speech data can be justified by following any type of utilitarianism based on the maximization of utility, for deontological ethics , using personal speech data is equivalent to a case of reducing humans to some instruments for some other purpose. The deontological ethics [3] will not justify the very activity of the collection of personal speech data, even if (a) the intended purpose of the collection of personal speech data might maximize the overall utility and, (b) the collection of personal speech data has consent from whom the speech is collected. The deontological approach never justifies any act that regards humans as mere means or instruments for any other purpose. Mostly, the utilitarian framework will be useful for developing an ethical framework for the domain of speech processing. Another problem to be noted here is that the mere focus on the utility aspect might deprive one to look at the different impacts of what the very speech recognition activity does to at least a moral agent. Consider the privacy-related issues in the input-confined ethical issues. The solution mostly depends on getting consent from the whom the speech data is collected. But then, to get the consent the entire teleological path needs to be explained from whom the consent is sought. Or else, the consent itself will not serve the purpose for which the consent is sought. The immediate purpose of the collection of speech data might be indicated to get consent. But the storage and transfer of such speech data may not ensure the entire path of the teleology, in the sense, the speech data that is being stored and transferred at a later period may have different purposes in a different context. In output-confined ethical issues, one can think of the very intrinsic nature of the releasing of processed speech using a deontological approach. Is there anything intrinsically wrong in releasing processed speech that will have an impact on at least a moral agent? Or is it merely instrumentally wrong? Such questions will

require a deontological framework. One could also consider the virtue ethical frameworks. In virtue ethics framework could the processing be a virtuous act by a system? Can the speech processing systems be virtuous? [5]

VIII. Immediate and distant ethical effects and the illocutionary force-content distinction

We make a distinction between immediate and distant ethical effects, as it was mentioned before that the data has the possibility of being stored and transferred at a later period in a different context. We distinguish between the immediate effect and the distant kind of effect that a released speech can have on a moral agent. The ethical issues need to be segregated according to the long-term and short-term effects that the released processed speech can have upon a moral agent. The collection of illocutionary force as opposed to illocutionary content will have a stronger effect on the nature of humans in the distant time than immediately. As there is essential relation between the features of the illocutionary force and the speaker, both the ethics of collection of the speech data and the ethics of contextualizing the speech data through some speech processing technique will have a distant ethical effect apart from the immediate ethical effect. One sort of distant moral issue would be the frequent effect that processed speech has on an individual for a long duration and the kind of alterations that this phenomenon would have upon the fundamental nature of humans. The distant ethical issues that arise out of the frequent dependency of humans on processed speech is to be distinguished from the distant ethical issue that arises out of data being used as information in a different context.

The epistemic threat that the deepfake technology creates on society is in fact a concern that exclusively falls as an ethical issue having a distant ethical effect. The use of deepfake technologies reduces the amount of information that it carries to the consumers who are the moral agents [6]. Much of these studies focus on the aspects relating to photography and videos. However, similar problems will emerge from the deepfake technologies that are of speech. One of the main epistemic threats that deepfake technologies give rise to is that it creates a situation where people frequently could end up in forming false beliefs. If what is intended by a particular deepfake speech is deception, people might still believe that what they hear is a genuine speech by someone.

IX. Speaker Dependent Ethical Issues and Speaker Independent Ethical Issues and the illocutionary force-content distinction

We need to distinguish between the ethical issues that arise from the speaker-independent speech processing system and the ethical issues that arise from the speaker-dependent speech processing system. There is a difference between (a) the collected speech after processing influencing a moral agent and (b) the input independent speech being released from the speech processing system influencing a moral agent. In the second scenario, speech processing is more independent since speech itself is the production of a processing system without depending on humans. Speaker Independent Ethical Issues arise if they result from speeches that are not spoken by any given human. Speaker Dependent Ethical Issues arise if they result from speeches that are spoken by any given human. The reason for the Speaker's Independent Ethical Issues relates to the notion of illocutionary force of the speech act. Consider the different features of the illocutionary force that emerge from the Speaker Independent speech technologies. None of those features of the illocutionary force have any actual speaker as a source. It is a case of having a speech act without a speaker. However, the speech act that emerges from the technology has the features of illocutionary force, but no human possesses any of the features of the illocutionary force. There will not be any kind of input-related ethical issue as no data is collected. How ethical is it to release a speech act whose illocutionary force is nothing but a creation? In this case, neither the ethics of data collection nor the ethics of contextualization have a great role to play. The issue with the speaker independent speech technology concerns the ethics of releasing of the speech. In this case the released speech functions as a speech act. More than the content of the speech act, the illocutionary force remains as the reason for the ethical issues. The speech act even if the human is not the source, is expected to occur with a state of the attitude of sincerity or the strength of the sincerity or the illocutionary point, etc. The illocutionary force whose sole source is technology influencing the human is the concern here. Here, again the distinction between short-term and long-term effects of the speaker independent speech is to be made.

X. Conclusion

In this paper, we introduced some concepts that are relevant for developing an ethical framework for the phenomenon of speech processing. Towards this purpose, we emphasized different stagewise ethical issues that arise in the case of speech processing. We distinguished between the input, output, and processing confined ethical issues. Using the speech act theory from Philosophy of Language, we showed the features of the speech data that are relevant to the study relating to the ethics of speech processing. Illocutionary force-content distinction clarified (a) the nature of ethical issues that arise in the case of speech processing, (b) the distinctness or uniqueness of the ethical issues relating to the speech data and (c) also showed the stagewise ethical issues arising in the case of speech processing. Using illocutionary force-content distinction we could also show the immediate and distant kind of ethical issues and the nature of ethical issues that arise from the Speaker Independent speech technologies.

References

[1] Alphago. https://www.deepmind.com/research/highlighted-research/alphago

[2] Adriaans, P.: Information. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy (2020)

[3] Alexander, Larry, M.M.: Michael. deontological ethics. In: Metz, T. (ed.) Stanford Encyclopedia of Philosophy (2021)

[4] Austin, J.L.: How to do things with words. Oxford University Press (1962)

[5] Constantinescu, M., Crisp, R.: Can robotic ai systems be virtuous and why does this matter? International Journal of Social Robotics **14**, 1547–1557 (2022)
[6] Fallis, D.: The epistemic threat of deepfakes. Philosophy Technology **34**, 623–643 (12 2021)
[7] Green, M.: Speech Acts", The Stanford Encyclopedia of Philosophy (2021)
[8] Grice, H.P.: Meaning. Philosophical Review **66**(3), 377–388 (1957)
[9] Jarosciak, J.: Social and ethical concerns of smart voice-enabled wireless speakers (2017)
[10] Konurbaev, M.: The Ontology of Speech and the Nature of Foregrounding, pp. 55–112 (02 2018)
[11] McLaughlin, E.C., Keith Allen, C.: Alexa, can you help with this murder case? (2017)
[12] Mill, J.S.: Ethical theory : an anthology / edited by Russell Shafer-Landau. 2nd ed. edn. (2013)
[13] Rios-Urrego, C., Vasquez, J., Orozco-Arroyave, J., Noeth, E.: Is There Any Additional Information in a Neural Network Trained for Pathological Speech Classification?, pp. 435–447 (08 2021)
[14] Ruiter, A.: The distinct wrong of deepfakes. Philosophy Technology **34**, 1311–1332 (12 2021)
[15] Searle, J.R., Vanderveken, D.: Foundations of Illocutionary Logic. Cambridge, England: Cambridge University Press (1985)
[16] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Driessche, G., Graepel, T., Hassabis, D.: Mastering the game of go without human knowledge. Nature **550**, 354–359 (10 2017)
[17] Smart, J.J.C.: Extreme and Restricted Utilitarianism1. The Philosophical Quarterly **6**(25), 344–354 (10 1956)
[18] Steunebrink, B., Wang, P., Goertzel, B.: Artificial General Intelligence: 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016, Proceedings (01 2016)
[19] Tiribelli, S.: The ai ethics principle of autonomy in health recommender systems. Argumenta **16**, 1–18 (2023)