

# チーム実習 レポート

学籍番号： 1F10220127

氏名： 竹本 志恩

指導教員名： 神場知成 教授

2025 年 1 月

東洋大学情報連携学部情報連携学科

## 論文要旨

近年、機械学習分野の研究は非常に盛んである。その中で特に大規模言語モデル (LLM) は、人の仕事を代替できるなど (要検討) の理由から注目を集めている。他方、日本における教育の変革が始まっている。2020 年以降、学習指導要領の改定が実施され、多くの学校が対応に追われている。その中核に”主体的な学習”があり、学生が自身から学ぶ姿勢を身に着けることで、激動の時代を生き抜ける力が付くという。その習得に貢献する方法として、文部科学省はアクティブラーニングを検討、議論していた。アクティブラーニングについて、特に複数人で課題に取り組むグループワークは、複数の研究からその有効性を示唆されているが (???)、一方で課題も多くある (???)。我々はこの課題の内、教師からの評価や生徒の参加意欲に着目し、LLM によるフィードバックでそれらの改善を図るアプリケーション、???を開発した。開発段階では、グループワーク中のディベート?を想定した評価実験を行い、2 名の参加者から構想に関して肯定的な意見をいただいた。しかし、フィードバックの詳細や UI については課題が残った。今後はそれらを改善し、より洗練されたシステムを目指す。

## 目次

1.	プロジェクトの背景と目的	1
1.1.	背景	1
1.1.1.	LLM と生成 AI の登場	1
1.1.2.	教育の社会課題	1
1.2.	目的	2
2.	先行研究	3
2.1.	アクティブラーニング	3
2.2.	グループワークの課題	4
3.	成果	5
3.1.	概要	5
3.2.	機能	5
3.2.1.	議事録作成	5
3.2.2.	評価作成	6
3.3.	成果物の懸念	6
3.4.	実験と結果	6
4.	貢献	8
4.1.	個人サーベイ	8
4.2.	チーム活動全体	8
4.3.	チームサーベイ 2024/6/12	9
4.4.	テーマ決定	9
4.4.1.	2024/6/19	9
4.4.2.	2024/6/26	10
4.5.	発表の準備	10
4.5.1.	2024/7/3	10
4.5.2.	2024/7/10	11
4.6.	開発	11
4.6.1.	7 月末-8/15	11
4.6.2.	8/16-9/15	12

4.6.3. 9月中旬-10月初頭 . . . . .	13
4.6.4. 10月-11月初頭 . . . . .	14
4.6.5. 11月 . . . . .	14
4.6.6. 12月 . . . . .	14
4.6.7. 1月 . . . . .	15
5. 振り返り . . . . .	16
5.1. タスク管理 . . . . .	16
5.2. サーベイ . . . . .	17
謝辞 . . . . .	18
参考文献 . . . . .	19
付録 A 評価に利用したコード例 . . . . .	20

## 1. プロジェクトの背景と目的

### 1.1. 背景

#### 1.1.1. LLM と生成 AI の登場

ChatGPT の登場以降、LLM による生成 AI は非常に盛り上がりを見せている。LLM は機械によって人の思考を再現しようとする機械学習や人工知能に関する研究の一つで、計算機の性能向上や情報の爆発的な増加を背景に研究が進み、実現にこぎつけた。この LLM にユーザが文字を入力するインターフェースを付け、入力に対する出力を得られるようにしたのが、GPT モデルを利用した OpenAI の ChatGPT である。この ChatGPT は従来の AI では難しかった、文章の意味や論理を理解したかのような、人間らしい文章を作成できる。つまり、人間レベルの推論が再現できる可能性があり、従来の仕事の代替や、より効率的な仕事の達成に役立つことが期待される。

ところで、生成 AI の生成対象は概ね 4 つに分類される。具体的には文章や音声、画像、動画である。以下の図 (図, 自作) に示すように、それぞれ調査時点で著名なサービスが存在している。(出典不要か)

また、野村総合研究所の調査によると、詳細な活用方法や利用サービスは分からないが、ビジネスに関わる人々の間でも活用が進められているという (出典)。次の図 (図, 出典) から、急速に認知度が高まっていることが読み取れる。

さらに、生成 AI を用いたサービスの開発、提供も活発に行われている。次の図 (図, 出典) は生成 AI に関する事業を行っている企業が示されている。このように、国内向けの事業だけでも相当数存在しており、それだけの勢いがあると分かる。

#### 1.1.2. 教育の社会課題

ここからは教育に関する課題に触れる。2020 年から 2022 年にかけて、新学習指導要領が小学校から中学校、高等学校の順に実施され、各段階で必修科目の増加や科目自体の再編など、多くの改定がなされた (出典?)。

これを受けて、学校や所属する教員は対応を迫られるが、以前からあった人手不足や一人あたりの労働時間の多さなど、相互に影響しあう問題に加えて、負担が増える形となっている (出典 教師不足に関する実態調査)。

私たちが注目したのは新学習指導要領の中核に据えられている‘主体的な学習’である。これは生徒が自分の力で考え、物事に取り組む学習を指しており、こういった姿勢を身に着けることが今後の社会で生き抜く上で重要だと文部科学省の資料に述べられている(出典)。この姿勢の習得のために、アクティブラーニングと称される手法がある(出典)。アクティブラーニングの一例として、グループディスカッションがある。生徒達が協同してある課題について話し合い、解決策を考えるが、こういった過程を通じて主体的な学習の姿勢が育まれるという(出典)。

## 1.2. 目的

私たちはアクティブラーニングの一種であるグループディスカッションに着目した。その手法に存在する課題を解決し、体験の質を向上することで、より主体的な学習姿勢の獲得を容易にすることが目的である。なお、課題については2節で触れる。

## 2. 先行研究

ここではアクティブラーニングとその課題について、先行研究を紹介する。

### 2.1. アクティブラーニング

[1]によると、アクティブラーニングは従来の講義型の学習とは異なり、教師からの情報をただ聞くだけでなく、理解を深めるという。ここで重要なのは、生徒自ら情報を理解し活用する点である。この手法は理解や利用の過程から、従来より知識が定着しやすい。加えて知識について考えることから、新たな学習機会の創造、ひいては主体的な学習姿勢に繋がると考えられる。また [1] ではアクティブラーニングの実現法は多様で、グループディスカッションなどの特殊な授業形態でなくとも可能だとされている。実際に [1] の表 2 では多くの方法が提案されている他、私もディスカッション以外の方法を体験したことが何度かある。例えば地理の授業で、5 分ほどの時間で Kahoot(<https://kahoot.com/>) を利用した方法が実施されていた。Kahoot はオンライン上で早押しの問題に回答するサービスで、この時は教師が設定した地理に関する問いを、より早く回答するべく生徒間で盛り上がりを見せていた。これは問いの内容にもよると思うが、少ない時間で学習の動機を与えられ、主体的な姿勢を身に着けるために有用な方法だと実感している。その一方で、ディスカッションも何度か体験している。一例として現代社会、今という公共の授業で行われたものを紹介する。それらは教室を賛成派と反対派の 2 つに分け、調査や検討の時間を設けた上で、討論を進めた。テーマは当時の社会課題であり、具体的には死刑は廃止すべきか、といったものだった。これは非常に面白い試みで、私自身今なお覚えているほど印象深かったが、もう少し少人数で進めた方が、偏りなく学習の効果が得られただろう。

さて、私たちは 5-6 人による 30 分ほどの議論を想定してシステムを構築したが、これはグループディスカッションを前提としたためである。[1] は大規模な講義であっても実現可能な手法を前提として、アクティブラーニングにおいてグループワークは必須ではなく、手法には工夫の余地があると述べている。一方で、私たちは主体的な学習姿勢の定着だけでなく、グループワークを通じた生徒間の交流やコミュニケーションを重視している。本システムは [2] で述べられる課題を解決し、グループワークの実施及び選択を容易にする。

## 2.2. グループワークの課題

ここではグループワークに存在する課題を説明する。[2]において、私たちはフリーライダーの出現や評価の公平性に着目した。なお、ここでのフリーライダーは、自らの意見を出さず、議論の進行に関わらない参加者を指す。フリーライダーがグループに居ると、課題こそ完了するものの、グループ全体の意欲が低下し、また参加した生徒が不満を持ち、グループワークへの嫌悪を抱きかねない。また、フリーライダーの出現には当人のやる気のなさとは別に、コミュニケーションが苦手などによる消極的な参加が原因の場合もある。そして評価の公平性について、評価の基準や過程に対する評価の有無から、評価に不信感を抱く生徒が発生するという。これらはINIADの1年次に行う情報連携基礎実習や英語の講義などで発生しており、私含めたチームメンバが実際に体験していた。この経験から、生徒と教師に対してGPTによる評価を与える本システムを考案した。[2]では補助教材による解決法が模索されているが、私たちは生徒に対して議論のフィードバックを評価として与え、自覚と自省の動機及び機会を作り、議論の過程をつまびらかにすることで公平な評価を志す。これらの手法から、課題の解決を目指す。



### 3. 成果

#### 3.1. 概要

私たちは議事録を作成し、それに基づいた評価を行うシステムを提案する。本システムは中学校、あるいは高等学校で実施されるグループディスカッションでの利用を想定している。また、5-6 人のグループが 5 つあると仮定し、図 4 のように、それぞれ指定された机の周囲に集まるとする。この時、机の上には学校から配布されたノートパソコンが設置され、そこから本システムを使用する。大まかな流れは次のようになる。まず各グループは通常通り議論を進め、その音声録音される。次に、録音した音声に対して Google の Speech-to-Text で音声認識を行い、文字起こす。最後に音声認識の結果を OpenAI の GPT-4o mini に渡し、評価を作成する。なお、作成した議事録と評価は生徒並びに教師に共有される。この評価によって、生徒は第三者からの議論のフィードバックが得られ、反省の機会が生まれる。反省を行うかどうかは生徒に委ねられるが、もし行う場合には、ディスカッションを重ねるごとに、通常の実施よりも主体的な取り組みの態度が養成されると思われる。また、教師は LLM の評価を参考に生徒へアドバイスを考えられる。実際には議論の過程が記録された議事録も提供されるため、より生徒の貢献度や態度を評価しやすくなると考えられる。

#### 3.2. 機能

次に、本システムの持つ機能を紹介する。

##### 3.2.1. 議事録作成

議事録作成について、録音と音声認識を行う。録音では、机の上に設置されたノートパソコンに備え付けの、指向性などを持たないマイクを利用する。その音声を一度 Google Cloud のストレージに送信し、Speech-to-Text によって音声認識を行った後、結果を txt ファイルとして保存する。この結果が議事録であり、本来は音声認識だけでなく話者分離も行う予定だったが、単語ごとに分離される上に精度が出ず、評価に差し障るため音声認識のみの適用となった。

### 3.2.2. 評価作成

評価は対象ごとに行った。1つは個人, もう1つはグループ全体である。それぞれプロンプト (付録 番号) を用意し, 出力を得た。

また, その前段階として誤字脱字の簡易的な修正も行っている。肝心の修正はあまり上手くいかなかったが, 句読点の付加や文ごとに区切るなど, 文の整備は成功し, LLM が解釈しやすくなったと思われる。

### 3.3. 成果物の懸念

以上がシステムの大まかな流れと機能になるが, 現時点でいくつか問題がある。まず Speech-to-Text について, プライバシーの不安がある。一度 Google Cloud に音声を送信する必要があり, 万が一通信の過程などで音声が漏洩すると, 生徒のプライバシーを保てない。

次に生徒に議事録や個人評価を共有する仕組みが出来上がっていない。これは当初実装する予定だったが, 開発に遅れが生じたため間に合わなかった。

また, 表示できる評価が最新の議事録に限定されている上, 具体的な評価ができない。これは開発中に行った実験のフィードバックで分かり, まだ実装できていない。実験については次の項で述べる。

### 3.4. 実験と結果

私たちは機能の開発後, 簡単な実験を行い, システムの使いやすさや, 評価の内容が目的に即しているかを確認した。実験は殿村が中心になって行った。状況は概ね想定通りで, 周囲に会話があって雑音が入る環境で, 机の上に置いたノートパソコンのマイクによって録音を行い, その評価を参加者に渡し, 質問に解答してもらう形でシステムの評価を行った。質問の内容や解答結果は付録 (Google form を全部張り付ける) を参照していただきたい。

実験の結果, デザインやアイデアの面では肯定的な評価だったが, 評価の内容や表示の仕方について, 批判的な評価を受けた。

特に評価の内容について, 実際の発言に触れず, 具体性に欠ける, といった意見が見られた。

そこで、それらを踏まえてプロンプトを見直し、再度評価を出力した。この際、Speech-to-Text と別に、WhisperX という Whisper に話者分離の機能を加えたパッケージを利用した出力も提供し、評価してもらった。これはおよそ話者分離をした場合に出てくるフィードバックについても意見や評価が欲しかったためである。

また、アンケートの内容に議事録に関する項目を加えた。

その結果 Speech-to-Text による議事録に関して、精度が低かったり、話者が分からないため振り返りに役立ちづらいという意見が出た。個人に関する評価については、以前よりも具体的な内容が加わり改善されているとした上で、評価の根拠を具体的に教えてほしい、といった透明性に関する要請があった。

これらの結果から、音声認識や話者分離の精度が十分にあれば、反省に役立つ議事録が生成できる可能性が示唆された。ただ、それらの精度が十分にあったとしても、評価に具体性や透明性が必要だと分かった。

## 4. 貢献

ここでは私のチームに対する貢献について、プロジェクトを段階ごとに分けて説明する。内容としては、私の取った行動とその理由や目的、本プロジェクトから得た学びや課題などの反省などが含まれる。

### 4.1. 個人サーベイ

チーム結成の前段階として、個々人で生成 AI に関するサーベイを行い、グループ内で発表した。私は最低限の知識の習得と、方向性の確立を目指して、主に生成 AI の基礎知識と利活用を中心に調べた。当初は活用の題材として、心理カウンセリングや詐欺広告の検知、リスクリングなどを考えていた。リスクリングは社会人が新しい知識や技術を学びなおすことを言い、実際に本学部や設置されている大学院では、社会人の受け入れや学習支援を行っている。ただ、このサーベイの内容は量、質ともに不十分であったと思う。活用事例として、複数の分野を見据えることは良いが、それぞれが手薄になるのはいただけない。それに加えて、生成 AI に関する技術やその動向にはあまり触れなかったことも反省点になる。これらは秋学期のプレ卒論で多少活かされるも、そちらでも不十分なサーベイのまま進めてしまった。研究に限らず、多くの事業で必要になる大切な段階だと思うので、今後は小規模なプロジェクトで練習を行いたい。

### 4.2. チーム活動全体

以降はチームの活動とそこでの貢献を説明するが、その前にそれぞれの役割に触れる。中間発表までは毎回何らかの議論を行っており、そこでは議事録を私か北が作成し、アイデア出しを佐藤、私、諸角が、進行や取りまとめを殿村が行った。議論に必要な情報は、主に北、私、たまに佐藤が事前に収集した。中間発表後は、それぞれ開発のタスクが振られていたため、毎回の議論は最低限の方針や現状の確認に留まった。開発全体において、機能の開発は私を中心に佐藤が補助し、API の検証は逆の分担で行った。Web アプリケーションのデザインや UI は諸角と北が、全体の総括や実験は殿村が担当した。

プロジェクト全体における私の貢献は、大まかには議論を生産的にするための議事録の作成や資料の作成・配布、発言による議論の盛り上げ、そして話者分離の実現に向けた取り組みだと思われる。

### 4.3. チームサーベイ 2024/6/12

個人サーベイの発表後、教育に関するテーマを扱う人でチームを組み、プロジェクトを開始した。まずプロジェクトで行う内容を決める必要があり、案を出すために教育の社会課題と、それと別に生成 AI に関する追加のサーベイを行った。ここで主な課題として挙げたのは、教員の人手不足である。教員の人数が不足し、一人の労働時間が増え、そういったデータから教員を敬遠する人が増え、人が足りないといった負の循環があるように見えた。こどもの全体数は年々減少しているが、教員も同様に減少しており、にもかかわらず業務の量は増えていて、負担が多い出典。こういった現状に対して、教員の業務自動化によって対応する解決策がいくつか見られた。内容としては、テストの作成や採点、苦手分野の解析とそれに応じた学習教材の作成などがあった。

6/12 はブレインストーミングを行い、図 (出典) のようにまとめた。ここから、人手不足に焦点を当てていった。議事録は付録 (番号) を参照していただきたい。

### 4.4. テーマ決定

#### 4.4.1. 2024/6/19

6/12 の議論を踏まえ、6/19 には一週間で調べたサーベイの内容を持ち寄り、それらと各々のアイデアをまとめ、議論の末にテーマを決めた。

分野として教育や福祉、介護があったが、最終的に議事録の提案が面白い、という意見で一致し、教育分野のグループワーク補助をテーマとした。この時、グループの作成も同時に行おうという話があり、プロジェクトの案がある人に対応したスキルを持つ人を割り振ったり、仲の良い人達を分割してコミュニケーションを促すといった案だったが、6/26 の議論の中で、時間的な懸念から消滅した。私としては補助を主軸に、他の案も生かす形でプロジェクトを進めて、皆の不満なくシステムを開発する想定でいたが、想像以上に時間が足りず、実現は難しかったと思われる。

こちらの議事録は付録 (番号) となる。

#### 4.4.2. 2024/6/26

6/26 にはテーマについてさらに議論し、どういった内容が いいかを考えた。グループワークの課題としては、教員が複数のグループに対して一人しかおらず、細部の把握や評価が難しい点、評価の客観性、負担の増加などが挙げられた。

ここで客観性については、恐らく透明性のことを指している。客観的で公平な評価をされていないと感じる生徒に対し、評価のフローや根拠を提示できれば納得が得られるはずで、それは評価の透明性の確保を意味する。プロジェクト終盤に行った実験においても、評価の透明性が課題として挙げられた。ここで十分に議論をし、前もって課題を把握するか、あるいは開発の遅れを最小限に留め、実験とその反映を何度もこなせる状況を作り出すべきだったと思う。

また技術的な課題も整理し、システムの大まかな動作や必要な技術についても議論した。ここで議事録の作成には音声認識だけでなく話者の認識も必要だと考え、実現方法を調べることになった。話者の識別は私から提案し、その後調べていったが、

それから生徒の評価について、どういった基準を設けるのかを話し合った。私は発言回数や量なら定量的に評価できてよいと考え、それらを提案したが、その場合議論に貢献しない会話が多い人も評価されるという指摘があった。では雑談かどうかを判定しようという方向で話が進み、場を暖める会話を誤判定しない必要がある、という意見にまとまった。この時は全く考えていなかったが、結果として、これらは評価で使ったプロンプトの中に組み込まれた。

こちらの議事録は付録(番号)となる。

#### 4.5. 発表の準備

夏休み前の発表に向けて、資料やシステムの内容を詰めていった。

##### 4.5.1. 2024/7/3

大まかな開発のスケジュールと、開発する機能を決めた。私はテーマの動機について、社会課題の解決として提案したが、それよりも評価の内容や API の選定などが優先され、十分に議論できなかった。この時は前回から引き続き、グループワークに関する資料を作成、

共有したと記憶しているが、今一度見るとどこから見ればいいのか分からない、雑多なメモ書きになってしまっている。私の中で十分に整理されないまま共有がなされ、それによってチーム全体の理解も曖昧になってしまったと思う。これはこの後も何度か行っていて、十分に反省すべき点である。

こちらの議事録は付録 番号 となる。

#### 4.5.2. 2024/7/10

主に発表資料に必要な要素の議論と、資料の作成を行った。ここで前回の懸念点だった教育に関する課題や解決策の動機について議論した。ただ、理解が曖昧だったためか、人手不足による教員の負担増加を解消する、という目標設定を行ってしまい、資料の整合性が取れなくなった。整合性は発表でも指摘されたが、最もその部分を理解していたはずの私に大きな責任があると思う。自身の持つ曖昧な理解を排し、その上で全員に説明を行って十分な理解を得る、というプロセスが必要だったのではないか。

こちらの議事録は付録 (番号) となる。

### 4.6. 開発

ここからは、私が開発で行った貢献と、そこで得た学びや課題、反省を述べる。

#### 4.6.1. 7月末-8/15

音声認識に用いる API や話者分離について調べ、試験的な実装を行った。

当初は API として、Whisper と Amivoice を検討していた。これらは料金が安く、ドキュメントも充実しており、利用が容易に見えたためである。そこで私が Whisper を、佐藤が Amivoice をそれぞれ担当し、音声認識と出来れば話者分離までを試して、上手くいきそうな方で開発を進める方針とした。

Whisper 単独だと音声認識は出来ても話者ごとに会話を分離することが難しい。調べたところ、Pyannote というライブラリを利用した例が多数あり、また WhisperX という、話者分離を行うために Whisper に諸々を加えたパッケージも存在した。

私は WhisperX をローカルで、Whisper + Pyannote の組み合わせを Google Colab にて検証した。前者は今なら CPU で動かす設定も出来たと思うが、当時は WhisperX のファイ

ル構成や GPU の使い方を理解しようとせず、GPU がいないため動作しないと結論付けて検証を中断した。後者はメモリ効率が悪いのか、実行途中で GPU の利用上限に達して動作しなかったと記憶している。その数日後に GPU マシンを手に入れ、検証できるようになったが、既に 8 月の半ばであまり時間に余裕がなく、渋々 Amivoice による開発に切り替えた。

当時の進捗は Slack に残っており、以下のようになっている。

8/8-8/9: 音声認識及び話者分離について調査, 考察 8/11: WhisperX の検証 8/12: WhisperX の検証続き, Whisper + Pyannote による音声認識 + 話者分離の検証 8/13: Whisper + Pyannote の検証 続き 資源枯渇により進まず 8/14: 個人的に使用するための計算サーバ購入, OS インストール, 設定 8/15: 設定/Whisper 検証続き

このように、1 週間で検証自体は完了しているものの、夏休みに入って一週間ほどは何もしていないか、ゆっくりと調べていたせいで、開発に遅れが生じてしまった。その間にもアプリケーションのインターフェースやデザインの開発は進んではいたが、中核となる機能開発が進まないと実験や改修に至らないため、かなり致命的な遅れが発生していたと考えられる。

春学期の講義が終了した後、毎週水曜日に欠かさずミーティングを行ったが、それでもこの一週間の遅れが発生してしまった。実は事前に組んだスケジュールは私が草案を作っている。発案者の私が守れなかったことはとても無責任であるし、それが原因の 1 つとなって話者分離の実装が叶わず、実験も一度のみ、しかも小規模なものになってしまった。この後の 9 月から 11 月にかけて、開発が思うように進められていないが、これは秋学期の講義が大変だったからである。長期的な目標の達成を見据えて、目の前のタスクの章かに真剣に向き合うべきだった。

#### 4.6.2. 8/16-9/15

佐藤が調べて多少動作を確認した Amivoice に焦点を当て、必要な機能の開発に従事した。この期間に完成したのは録音機能のみで、ほとんどの時間を Amivoice による音声認識に費やした。

当初は Websocket で API のエンドポイントと通信し、リアルタイムで文字に起こす予定だったが、認識精度があまりよくないことと、実際にコードを書き始めると大変だったことから、非同期の HTTP に通信方法を変更した。Websocket 自体は講義で触れており、しかも公式の用意したソースコードなどがあったため実装を楽観視していた。しかしいくつかの



問題が生じ、時間の兼ね合いで変更になった。記録がないため憶測になるが、公式の用意したライブラリ <https://github.com/advanced-media-inc/amivoice-api-client-library> の検証済みバージョンが Python3.7.4 までしかなかったことと、Websocket の書き方が良く分からず、どうやってシステムに組み込めばいいか分からなかったことが原因だったと思われる。

非同期 HTTP による通信に変更後、API による音声認識の実装は一旦佐藤に任せ、私は音声の録音機能を開発した。この際、Web ページ上にあるボタンから Python ファイルを動かすのに Ajax (出典) を使ったが、エラーやバグのない動作を実現するのに非常に手間取った。ここでエラーハンドリングが以前よりずっとマシになったほど、原因の良く分からないエラーが何度も出た。一連の実装で、HTTP などのネットワーク通信はエラーをきちんと扱う、と改めて認識させられた。

#### 4.6.3. 9月中旬-10月初頭

佐藤が音声認識のひな形を作っていたので、それを元に音声認識及び話者分離の機能を実装した。ひな形にエラーが残っていたため解決し、その後 Amivoice のドキュメントを見ながら実装を進めた。API 使用時に特定の要素を与えると話者分離が可能になるため、これを用いて一連の機能の実現を目指した。

実装について、話者分離の結果が Json ファイルとして渡されるのだが、その構造の把握に時間がかかった。ドキュメントに書かれている構造を丁寧に読み解いて、それを確認することを繰り返し、何とか実装が完了した。一連の Amivoice の実装を通して、ドキュメントを読んでコードを書き、理解を深めるという流れが如何に大切かが良く分かった。この学びは秋学期の講義や、Speech-to-Text, WhisperX の利用でも活きた。

また、検証として、Youtube から取得した 30 分ほどの音声を入力し、目視で確認した。今回は WER(Word Error Rate) や CER(Character Error Rate) などの、誤り率を評価指標として算出しておらず、そういった数値による性能の可視化は今後の課題だと考える。確認の結果、音声認識自体はあまり間違いはなく、意味が取れる文章になっていたが、話者分離に関してはまるで上手くいかなかった。話者の人数の下限や上限を変更したり、15 分ほどの音声を用いて検証しても同様の状態が確認された。もしかすると精度向上の手立てはあるのかもしれないが、その時は Amivoice の話者分離の精度は悪いという認識がチームで共有された。

認識精度: <https://www.ai-shift.co.jp/techblog/4512>

#### 4.6.4. 10月-11月初頭

精度に不安のあるものの話者分離まで実装が完了したため、私は議事録の評価機能の作成を開始した。幸い、GPT の利用は先の経験が活きて、そこまで時間がかからずに完了した。しかし、その後中間試験や大学祭があって、チーム全体で思うように活動が進められなかった。ここはしっかりとタスクを切って、着実に進めるべきだったと思う。

#### 4.6.5. 11月

諸々が終わりようやく活動を再開する。実装は、Django の model 機能を使い、データベースを Python から利用するコードを書こうとして、時間を費やした。私の実装だと、データベースのデータ入出力に Python を用いている。そのため view から Python で定義した関数にアクセスし、諸々の処理を行ってデータを渡してもらうのだが、以前から使っている Ajax など、複数のエラーが生じ、解決に時間がかかった。

まず、録音の途中でシステムが止まるようになってしまった。これはプレ卒論で試行錯誤していた時、WSL2 の Ubuntu を一度再インストールしており、それによって音に関するパッケージがなくなったことが原因だと思われる。解消にはそれなりの時間を費やしたが、以前録音の実装中に既に Ubuntu、特に WSL の環境に起因する音関係の不具合に遭遇していた、そのおかげで無事に解決した。

次に jquery のバージョンによってエラーが生じた。これは調査を含めて二日ほどかかったが、無事に原因を特定し、解決した。

問題は Ajax による送受信で、これが私の思うように動作せず、またプレ卒論などのタスクに時間がかかり、結果として実装に 2 週間近くかかってしまった。

私は夏休み中に川原先生の下で研究の補助を行っていて、その関係で論文を書き、12 月には研究会に参加する予定だった。この論文執筆のタスクを甘く見積もっていたせいで、思うように進められず、チームに迷惑をかけてしまった。

#### 4.6.6. 12月

以前から話に出ていた音声認識及び話者分離の手法変更について、佐藤が各種サービスを検討した結果、Google の Speech-to-text が話者分離もそこそこの精度が出て良いという

ので、システムへの実装は私が行うことにした。変更の目的は話者分離の精度向上である。佐藤の作ったコードと公式のドキュメントから実装はすぐに完了し、何度か検証を行った。しかし、そこまで分離性能はよくないように見えた。そこで、一層の精度向上を図るべく、以前検討していた Whisper を活用した話者分離をもう一度取り組む決心をして数日取り掛かったが、研究会に参加する関係で中断し、再開は年明けの最終発表前となった。

この他に実験のための環境の整備や軽微なバグの修正と、GPT による誤字脱字修正を目指したプロンプトの作成、検証を行っていた。

#### 4.6.7. 1 月

年が明けてしばらく経ったころ、WhisperX を GPU マシン上で動かし、実験で得られた音声の話者分離済みテキストを入手。それを使って評価を出力し、殿村を経由してシステムの評価を再度行ってもらった。

また、評価の表示が期待した内容と異なっていたため、表示を行う Python ファイルを修正した。

年末年始は他の課題をやっていて、講義が再開してからはプレ卒論の追試を進めていたため、この他に目立った活動はしていない。

## 5. 振り返り

プロジェクト全体を見返すと、総じて音源分離に振り回されていたように思う。そして、学ぶことも多かったが、それ以上に反省すべき点が多かった。実は議事録の作成と評価、そして音源分離の活用は私が提案しており、他にも実現性を無視して、各自が多くの時間を割くことを前提とした提案を何度か行っている。今後の活動では限られたリソースを活用して実現することを意識したい。

また、それに関連してタスクの管理と、サーベイについて反省がある。それぞれ以下で述べる。

### 5.1. タスク管理

中間発表前後までは一週間以内に行うタスクと目標を各自設定し、その達成を目指して動いていた。しかし、それ自体があまり丁寧ではなかったし、途中からはあまり意識できていない。

具体的にどう丁寧でないかという点、2つ特徴がある。

1つは切り分けと割り振りが上手くできていないことだ。ある目標を設定するのはいいが、その達成に向けて大まかなタスクの設定やチーム内での共有ができていない。特に私の場合、仮にそれらを行ったとしても、タスクを一日で終わる量に切り分けられていなかった。そして機能の開発に限って言えば、私、佐藤、殿村は上手くタスクを振り分けられておらず、誰かが暇を持て余す状態が長く続いた。

またチーム全体として、それぞれのタスクを概要だけ把握している状態だったが、誰も指摘をすることなく開発が進んだ。確かに、デザインやUI、機能、検証、サーベイといった漠然としたタスクと、その週の目標だけ共有していたが、これでは日々の進捗があまりに不明瞭だった。実際に誰かが遅れたり、何か困っているとしても、本人が皆に報告するまで気づかないことが何度かあった。

これらが音声認識や話者分離の開発が遅くなり、他の開発も遅々とした進みになった原因だと私は考える。もちろんきちんとタスクを処理しなかった私に大きな責はある。しかし、一度目の失敗の時に、チーム全体で日課として進捗を共有したり、タスクを分割して暇な人に任せるといった工夫を行えば、より効率的に開発が進められていたし、チーム開発の経験値も増していただろう。

## 5.2. サーベイ

本プロジェクト全体について、サーベイが圧倒的に不足していた。サーベイの量、質ともに全く十分ではなかったと思う。

最初のサーベイの時点で内容が怪しかったが、それ以降もあまり改善されることなく進んでいった。特にプロジェクトの背景と目的について、根拠が弱い上に、アクティブラーニングの意義やその具体的な手法への理解がチーム全体として不足していた。私はそれを懸念だと考えていたが、それを各員に強制するのは一講義の領域を超えろと思い進言しなかった。その結果、サーベイの不足を補うため、開発で手の空いた佐藤が延々と文献を読み漁り、結果何も残さないという惨事を引き起こしてしまった。実際、本レポートの先行研究の項は佐藤が見つけた文献を含んではいるが、私なりにまとめた内容が書かれている。また開発についても、学術的な成果を根拠にした機能はない。真にチームメンバやプロジェクトのことを想うのであれば、少なくとも殿村、佐藤には協力を呼びかけてサーベイを共に行ったり、その成果を簡潔にまとめたレポートを全員で共有して、認識を統一した方が良かったのではないかと。それにその方が、より問題の解決に効果的なシステムを開発できたと思う。

謝辭

## 参考文献

- [1] 泉美貴, 小林直人. アクティブ・ラーニングとは (総論). 薬学教育, Vol. 3, , 2019.
- [2] 碓山恵子, 木村尚仁, 細川和彦, 塚越久美子, 菊池明泰. 3g16 グループワークの課題を改善し深い学びを促進する補助教材の検討. 工学教育研究講演会講演論文集 第 65 回年次大会 (平成 29 年度), pp. 566-567. 公益社団法人 日本工学教育協会, 2017.

## 付録 A 評価に利用したコード例