

プロジェクト名：
申請者名：藤巻晴葵

概要

単一のハードウェアでユーザーの対面、電子的なコミュニケーションを理解し、ユーザーの継続的なコミュニケーション、創出を支援し、コミュニケーションコストを最小限に抑えることを目的としたソフトウェア。

1. 何を作るのか

背景

私は自身の経験を振り返ると、1人で作業を進めることを好む場面が多々あった。この傾向は、単独作業の方が効率が良いと感じてしまったからである。しかし、よく言われる「早く行きたければ1人でやれ、遠くに行きたければみんなでやれ」という言葉には、深く同意する。世に存在する大企業や革新的なサービスは、一人の力では生み出されていない。1人で大きいことを成し遂げることは難しい。複数人で行うことで達成することができるのだ。

私自身はこれまで課外授業やプログラミングのコンテスト、ハッカソンなどを通して、約10回ほど数人から10名の規模でのチーム開発を行ってきた。その中でチーム作業の価値を感じるが多々あったが、それでもなお、1人で作業を進めたいと感じる瞬間は少なくなかった。これは、1人であればより効率的に作業を進められると感じる場面が何度もあったからである。1人で行うことが効率的に思う理由はなぜだろうか。私がこのように感じる主な理由は、コミュニケーションにあると考える。

チームの全員が気持ちよく仕事をするためには、高い水準のマインドシェアを保つことが必要不可欠である。高い水準のマインドシェアを保つことは、チーム内での意識を統一し、同じ方向を向かせることである。チーム内での高い水準のマインドシェアを保つためには、密なコミュニケーションが必要になる。ここでの密なコミュニケーションというのは、達成したい目標についてわかりやすく説明するというだけでなく、各チームメンバーについて深く理解し、自分とメンバーとの関係を深めるということを指す。これを達成するとチーム全員それぞれが効果的な行動を行い、その行動全てが正の方向で進み、結果として、チーム全員が高い満足度で仕事を達成できる。逆に、このコミュニケーションが不十分な場合、チームの一体感が損なわれ、各メンバーが迷走することになる。私は、チームでの開発において、この高度なコミュニケーションとマインドシェアを実現することが、単独で作業をする際のコストパフォーマンスに勝ることが難しいと感じてきた。これは、チーム作業のパフォーマンスが個人のそれを上回るまでのギャップ、いわゆるキャズムが存在するためである。

このコミュニケーションのキャズムについて、私が経験してきたチーム開発のほとんどはこのキャズムが大きくなりやすい状況であったこともあり、私自身、このキャズムを超えることが困難であった。私が行ってきたプログラミングのコンテスト等でのチーム開発というのは、企業内で行う一般的なチーム開発とは違い、給料が発生するものではない。そのため、チームメンバーそれぞれのモチベーションがパフォーマンスに直結する。また、開発に取り組む理由や優先順位についてはメンバーごとによって大きく違う。これらにより、チームのパフォーマンスを高めるためには、メンバーそれぞれと密なコミュニケーションをとり、彼らを深く理解し、動機づけることが一般的なチーム開発よりも必要となる。

人を理解する、人との関係を深めるということにはコミュニケーションが必要不可欠であるが、人との関係値というのは、バイナリーで表すことはできず、グラデーションである。つまり、一朝一夕のコミュニケーションだけでは関係を深めることはできず、小さなやり取りの積み重ねが重要である。しかし、人との関係値はグラデーションのように変化するため、目に見えにくく、やり取り全てが正に働くかも不明確で、ゴールが見えにくい。

加えて、人との関係を効果的に深めるためには、良いタイミングと相手を思いやるコミュニケーションが必要になってくる。具体的には、相手と自分の立場関係、相手の年齢や嗜好などを考慮してコミュニケーションしたり、相手と何か対面で一緒にいることがあった、何かをしてもらった、相手に関する出来事があったときにコミュニケーションしたりすることは重要であり、効果的に関係を深めることができる。しかし、これは効果的ではあるが、考慮する点が多く、心理的ハードルが高い。

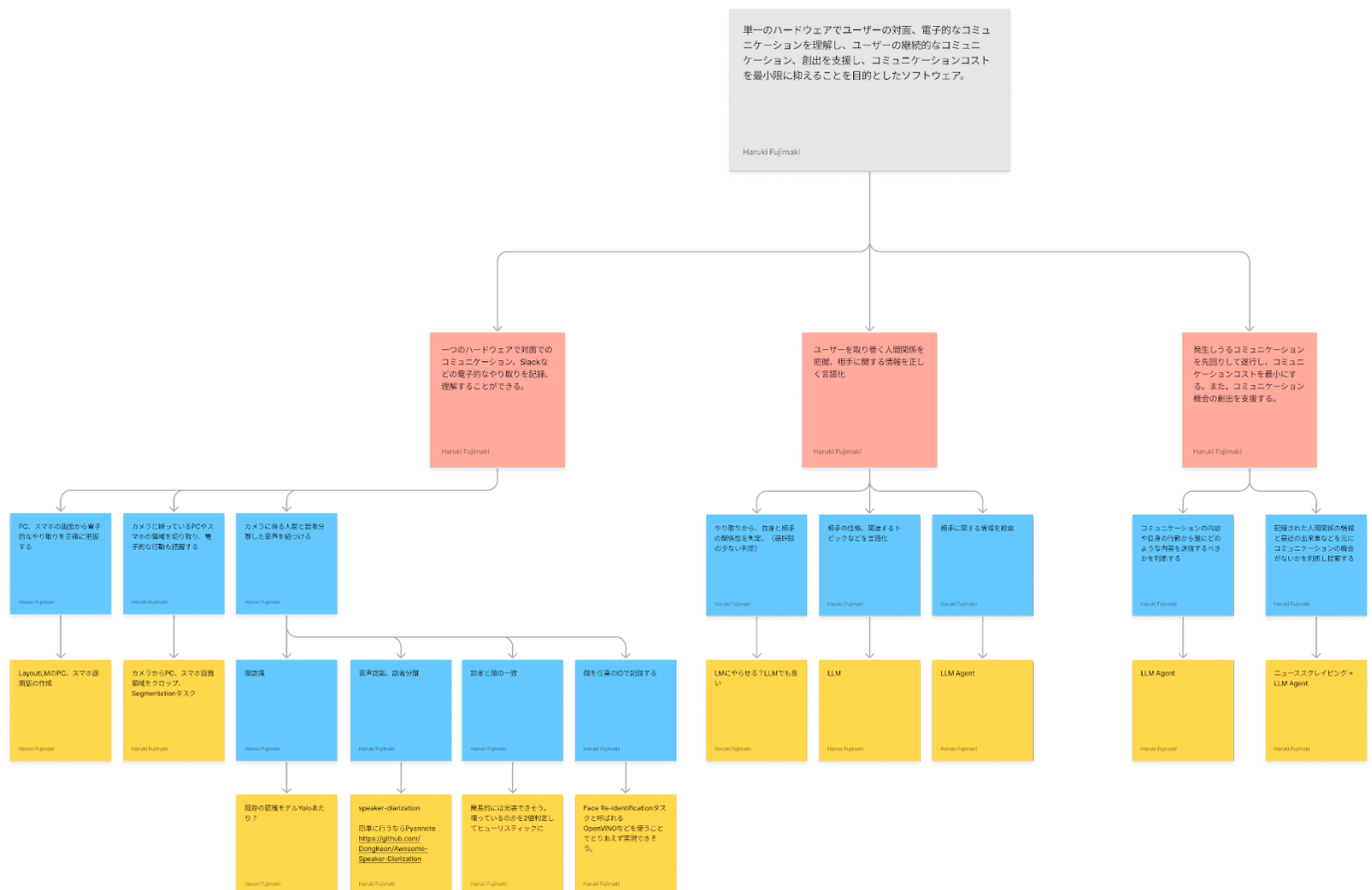
私は上記で述べた人との関係を深めることは一朝一夕では成し得ず、ゴールがなく、継続性が大切である点、相手を思いやる効果的なコミュニケーションというものは心理的ハードルが高いことから、しばしばコミュニケーションに対して、優先度を下げてしまいがちであり、結果として1人でやる方が効率がよいと思ってしまうことがあったのだ。

そして、私はこの認識を踏まえ、この1人でやったほうが良いというキャズムを乗り越え、自分を取り巻く人たちと深く関係を継続的に築くため、自分がコミュニケーションを取るという行動を起こす前に、コミュニケーションの行動を推奨・遂行してくれるシステムを開発し、コミュニケーションの心理的ハードルを下げ、効果的なコミュニケーションを継続して取れるようにしたいと考えた。このシステムは、ユーザーの行動を監視し、適切なタイミングでコミュニケーションを促すことにより、人との関係性を強化する。対面でのやり取りだけでなく、電子的なやり取りを含め、生活のあらゆるシーンでのコミュニケーションをサポートする。具体的には以下3点の機能を軸としてシステムを開発したいと考えている。

1. 一つのハードウェアで対面でのコミュニケーション、Slackなどの電子的なコミュニケーションを記録、理解する
2. ユーザーを取り巻く人間関係を把握、相手に関する情報を正しく言語化しDB管理
3. 発生しうるコミュニケーションを先回りして遂行し、コミュニケーションにおける心理的ハードルを最小にする。また、コミュニケーション機会の創出を支援

これにより、コミュニケーションにおける心理的ハードルを最小限に抑え、より効率的かつ効果的な人間関係を構築することが可能になる。私はこのシステムを通じて、チームでの開発だけでなく、より広義のコミュニケーションにおけるキャズムを埋めることを目指す。

提案するもの



一つのハードウェアで対面でのコミュニケーション、Slackなどの電子的なコミュニケーションを記録し、理解する

私は、自分を取り巻くあらゆる人とのコミュニケーションをより効率的かつ効果的に行いたいと考えている。そのためには、システムに対して自分が日頃から行っているあらゆるコミュニケーションについて、より詳細かつ幅広く理解させる必要がある。人とのコミュニケーションは対面でのやり取りやメッセージアプリによる電子的なテキストコミュニケーションなどが挙げられる。本システムでは、色々な形態のコミュニケーションをできる限り詳細に記録し、システムが理解できる仕組みを構築したいと考える。一般的に多岐に渡るコミュニケーションを記録しようとしたとき、その形態が増えれば増えるほど専用のハードウェアや処理を増やすことになる。例えば、電子的なやり取りを行うサービスだけに絞ってもSlackやDiscord、Messengerなどを複数プラットフォームを利用している。これらのプラットフォームそれぞれに対してAPI等を利用しやり取りを抽出することは正確ではあるが、汎用性がなく非効率である。また、スクリーンをキャプチャし、それを処理するという方法も全てのやり取りを記録しようとする自分が所持しているPCやスマホの端末に対して専用のソフトウェアを開発し、それをインストールさせる必要があり、これも非効率で汎用性がない。

そもそも、人とのコミュニケーションは自身の五感を通して行っているのだから、ユーザーに対してそれぞれの媒体でのコミュニケーションを記録するために追加の作業を要するのはナンセンスである。そこで、私は単一のカメラ、マイク、 α を搭載したハードウェアで対面のコミュニケーション、電子的なコミュニケーションを記録、理解できるシステムを開発する。

このシステムはハードウェアを通して取得した視覚情報、音声情報から、やり取りを記録する。対面での会話では顔認識と音声認識を行いやり取りを記録する。この時、認識した顔や音声に対して一意のIDを付与し、同人物との会話を継続的に記録できるようにする。また、複数人での会話についても話者分離を行い、やり取りを記録するようにする。電子的なやり取りについては視覚情報からPC、スマホの画面があった際にその領域を検出し、その部分をクロップし、専用の処理を行う。ここでの専用の処理として、私はPC、スマホ画面特化のマルチモーダル言語モデルを作成したいと考えている。具体的にはLayoutLMという文書画像と文書の文字情報、レイアウト情報を与えて文書の意味理解を目的とした言語モデルを応用し、PC、スマホ画面情報意味理解に特化した言語モデルを開発し、そのモデルを活用して電子的なやり取りの抽出を行いたい。GPT-4Vのような大規模マルチモーダル言語モデル（Vision LLM）を利用しない理由として、Vision LLMは抽象度の高い応対は得意であるが、細かな視覚的な応対は精度が低く、精度のチューニングが難しい。また、計算コストが高いため高い頻度で推論するには現実的ではない。専用の画面意味理解言語モデルを開発することによって、より正確な情報の抽出が行うことができ、計算コストについても抑えることができる。

以上をまとめると以下の3点が主に開発する内容である。

- カメラに映る人間と話者分離した音声を紐づける
- カメラに映っているPCやスマホの領域を切り取り、電子的な行動も把握する
- 専用のPC、スマホの画面情報意味理解言語モデルを作成し、電子的なやり取りを正確に把握する

ユーザーを取り巻く人間関係を把握、相手に関する情報を正しく言語化

- やり取りから、自身と相手の関係性を判定、（選択肢の少ない判定）
- 相手の性格、関連するトピックなどを言語化
- 相手に関する情報を検索

発生しうるコミュニケーションを先回りして遂行し、コミュニケーションコストを最小にする。また、コミュニケーション機会の創出を支援する。

- コミュニケーションの内容や自身の行動から誰にどのような内容を送信すべきかを判断する
- 記録された人間関係の情報と最近の出来事などを元にコミュニケーションの機会がないかを判断し提案する

未踏期間中の目標

未踏期間中の目標は以下の通りである。

- ほげほげ

2. どんな出し方を考えているか

私は、このシステムをマルチプラットフォームとして出すことを理想として考えている。PC、スマホから自分の行動データをアップロードすることができ、どの端末からでも自身のコミュニケーションを円滑化させることを可能にする。

3. 新鮮さの主張、期待される効果など

- 早く行きたいなら一人で行け、遠くへ行きたいならみんなで行けという有名な文章があるように、簡単なことというのは1人でもできるし、その方が早いケースも多い。
 - だが、何か大きなことをしたいという時には、1人ではできないことが多く、複数人で行うことでできるようになる。
- 大きいことは複数人でしかできない。だが、10人いたら10倍の速度で仕事ができるというのも嘘であり、それが10倍にならないのはコミュニケーションコストによるものである。
-

4. 具体的な進め方と予算

主な開発場所

筑波大学春日エリア 加藤研究室

使用する計算機環境

MacBook Pro, GCP, etc.

使用する言語・ツール

Python, JavaScript

開発計画

予算内訳

人件費

まず、提案者である藤巻は720 [時間] x 2000 [円]=1,440,000 [円]。また、開発の一部を他の学生などに委託する。この委託では時間で換算し、合計500 [時間]と考える。以上より人件費として2,440,000 [円]とする。

物品

まず、一つ目として、提案するシステムはいくつかの深層学習モデルを利用することを想定しているため、そのモデルを学習・推論するのに利用するGPUを搭載したクラウドコンピューティングのコストが発生すると考えられる。

二つ目として、本サービスをユーザーに利用できるような形で提供するにあたってのサーバーをホストするためのクラウドコンピューティングコストが発生する。

三つ目に、場合によっては本システムの専用ハードを開発する場合がある。そのためマイコンなどの購入が必要となる可能性がある。

以上3点の費用を見積もると以下ようになる。

- 深層学習モデルの学習・推論
 - 400,000 [円]
- サーバーホスト費用
 - 50,000 [円]
- 専用ハード開発
 - 150,000 [円]

人件費、物品費用を合計すると、上限の2,880,000 円を超過する。超過分は提案者の私費によって賄い開発を進める。よって、上限金額の2,880,000 円を申請する。

5. 提案者の腕前を証明できるもの

藤巻晴葵

現在に至るまで、授業、研究、ハッカソン・コンテスト、個人開発、会社で数十のシステムの開発を行ってきた。基本的にはソフトウェアの開発をメインとして行ってきたが、ハードウェアの開発に関してもいくつか作成してきた。また、高専4年次に会社を創業し、その会社での活動がいくつかのメディアで取り上げられている。以下にこれまでの成果を示す。

2019年度 全国高専プログラミングコンテスト 課題部門

- 最優秀賞・文部科学大臣賞
- ほげ

2020年度 全国高専プログラミングコンテスト 課題部門

- 最優秀賞・文部科学大臣賞

高専ディープラーニングコンテスト 2020

- 最優秀賞

情報処理学会 コンシューマ・デバイス&システム（CDS）研究会 CDS36

- 優秀発表賞

総務省 2022年度異能vation

- 異能ジェネレーションアワード受賞

6. プロジェクト遂行にあたっての特記事項

現在、藤巻は筑波大学情報学群知識情報・図書館学類3年に所属しており、来年度4年に進学する。藤巻は大学の授業をほとんど取り切っており、4年では卒業研究がメインとなる。研究室の指導教員から、未踏の応募について容認されており、プロジェクト遂行に当たって問題になる可能性は極めて低い。

7. IT以外の勉強、特技、生活、趣味など

藤巻晴葵

趣味はいくつかあるが、主に二つある。一つ目はサイクリングである。筑波大学へ編入する前の高専での在学中は家から高専まで、アップダウンのある片道10kmをクロスバイクで登校して

いた。そのため、自身の中でのクロスバイクでの移動は普通であり、体を動かしたくなった際には100km未満の距離のサイクリングをよく行なっている。二つ目はアニメ、ドラマ、映画などのPGC系動画コンテンツの視聴である。特にアニメは毎シーズン必ず何本かのアニメを見るようにしている。好きな内容の傾向としては、バトル系というよりも、心理描写がしっかりと描かれているヒューマンドラマなどを好みとしている。

また、大学のサークルなどを通してイベントの運営などを行なっている。TEDxUTsukubaという学生団体では、毎年100名規模のイベントを開催している。このイベントではワークショップや6名のスピーカーによる登壇セッションなどがあり、自身は協賛していただけるパートナーとの交渉や、翻訳システムの開発に携わった。また、エンジニアを対象としたミートアップ運営に携わっている。イベントを通じて、様々な背景を持った人たちと関わるができることに喜びを感じている。

8. 将来のITについて思うこと・期すること

近年、LLMが話題になったことによって、エンジニアの業務を改善するためのツールが多くリリースされ、普及も進んでいる。例として、CursorやGithub Copilotがあげられる。どちらもコーディングを支援するツールである。今までコードエディターに搭載されていたルールのコード補完とは違い、先ほどのツールはプログラマーの意図を高度に先読みし、ほとんどのコードを補完してくれる。私はこのツールの中でもGithub Copilotを利用しているが、体感として、自分がコードを書くという作業が1/2になったのではないかと感じている。それほどに強力なツールである。ここで一つ思うことがあるかもしれない。「このツールが発展していく、AIが発展していくことでエンジニアがいらなくなるのではないか」と。私は、エンジニアという存在がなくなることはないと感じている。一方、AIの進化によって、より開発することのコストは小さくなっていくと容易に想像がつく。今、あるサービスを開発するのに1年かかったとして、10年後には1週間、1日、1時間で作れるようになると考えられる。だが、考えてみてほしい。プログラミングに関しては、過去10年間の間で新しいプログラミング言語やフレームワーク、ライブラリの登場によってどんどんサービス開発等のコストが下がっていることに。しかし、今もプログラマーという需要は大きい。結論として、技術的なパラダイムシフトがあっても本質的に人間が行う活動というのはなくなるのではないのだ。確かにプログラマーという存在はなくなるかもしれない。だが、人間が思考するというのは自分のクローンが作られるまで続く。エンジニアという存在も時の流れによって、当時やっていたことをしなくなり、また違う部分に焦点を当てて活動するようになる。私は将来のITは日々新しいことに取り組み、それぞれが常に誰かに影響を及ぼすよりカオスな状態となると考える。