

社会的孤立の解消を目指した AI チャットシステムの開発

徳島大学大学院社会産業理工学研究部
松本 和幸

1. はじめに

現代社会において、多くの人々の日常生活においてストレスは避けられないものとなっており、ストレスを感じていると答える労働者の割合は年々増加傾向にある。しかし、仕事や職業生活の悩みなどを相談できる相手がいる労働者の割合と、実際に相談したことがある労働者の割合の間には差がある。身近で親しい人に相談できる相手がいる人は多いものの、相談できたとしても、日々ストレスを感じ続けているのが現状であり、専門家に相談するまでには至っていないことが推測できる。

また、近年、SNS によるテキストベースのコミュニケーションツールやメンタルヘルスケアアプリのサービスが普及し始めている。メンタルチェックを気軽に行えるサービスやストレスを自己管理し、セルフケアするツールの需要が拡大しているといえる。

以上のことを踏まえ、本研究では、ストレス管理 AI チャットシステムを独自に構築し、このシステムで収集したチャット履歴データをもとに、個人適応可能なストレス検出モデルを構築することで、低ストレス者と高ストレス者の判別を行うことを目的とする。低ストレス者は、本システムのストレス管理機能によってセルフメンタルケアを行い、日々のストレスを自覚し解消する。高ストレス者に対してはカウンセラーに誘導することでストレスを解消することを想定している。本報告では独自に構築したストレス管理 AI チャットシステムについて説明し、2024 年 8 月から 11 月に実施した実験で収集したデータをもとに、各ユーザのチャット履歴とストレスレベルを測るアンケートの分析結果をもとに構築したストレスレベル予測モデルについて述べる。

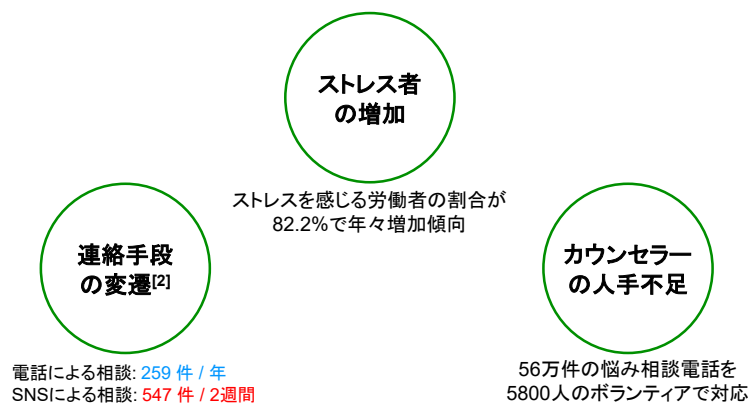


図 1. 研究背景

2. 先行研究

長野県と LINE 株式会社による LINE を活用したいじめ等相談の中間報告資料¹⁾によると、2017 年 8 月に「LINE を利用した子どものいじめ・自殺対策に関する連携協定」の一環とし

て、LINE を利用した中学生・高校生たちのいじめ・自殺等の相談を行った。その結果、9 月 10 日～23 日の 2 週間にわたり開設された LINE の相談専用アカウント「ひとりで悩まないで @長野」には、長野県内の中高生 390 人から計 547 件の相談が寄せられ、前年度 1 年間の電話相談 259 件を大きく上回った。このことから、テキストチャットを用いた相談に一定の需要があるといえる。しかし、文字中心のコミュニケーションでは、細かな感情の変化やニュアンスがとらえきれず、意思疎通に限界があるため、電話相談に切り替えて相談を継続する必要があることが課題として挙げられている。

メンタルヘルスケアを行うアプリとしては、うつ病に対する心理学的支援を目的としたスマートフォン向けのアプリが数多く開発されている²⁾。今後、メンタルヘルスケアを行うアプリの開発において、以下で述べる三点に特に注力する必要があると考えられる。第一に、セキュリティ対策を含むアプリの安全性を最優先事項として設計し、ユーザのプライバシーを確実に保護することである。第二に、アプリを通じて提供される医療情報の質と正確性を担保し、信頼できるエビデンスに基づいたコンテンツを提供することである。第三に、現状の抑うつ症状に対する心理学的支援アプリの治療法の種類が限定的であることを踏まえ、より包括的で多様な治療アプローチを組み込んだアプリの開発を目指すことである。さらに、臨床家との協力体制を構築し、アプリが実際の治療プロセスに効果的に統合できるよう設計することも重要な課題と考えられる。これらの点に留意し、臨床的有用性の高い、安全かつ効果的なうつ病患者支援アプリの開発が今後、発展を遂げていくことが期待される。

SELF アプリ³⁾ は、自然な会話で AI がユーザの生活を把握・理解し、メンタルケアやストレスケア、ライフログ、情報提案など、ユーザに個人適応するアプリケーションである。このアプリにはユーザのメンタルケアを支援する機能が多数実装されている。例えば、AI との会話のやり取りの中でユーザの特徴や長所・短所を性格分析し客観的に伝える機能や、ユーザの興味・関心を把握して、インターネット上のおすすめ記事を提案する機能などがある。AI との会話は提示された選択肢から選ぶ形式であるが、AI はユーザの選択した内容に応じて自然な会話を行うことができる。

3. ストレス管理チャットシステムの開発

本研究では、ストレス管理 AI チャットシステム (Stress Management AI Chat System, SMACS) を開発し、このシステムで収集した情報をもとに個人に適応できるストレス検出モデルの開発を目的とする。SMACS は、ストレスを抱える人々の支援を目的とした Web アプリケーションである (図 2)。ストレス検出モデルはシステムの利用データ (対話履歴やアンケート回答内容) をもとに機械学習アルゴリズムにより構築し、低ストレス者と高ストレス者を判別する機能として将来的にシステムへの実装を目指す。チャットシステムによる AI チャット機能は、OpenAI の ChatGPT の機能を API 経由で利用する。対話履歴やユーザのプロフィール情報をプロンプトに含めたカスタムプロンプトを動的に生成することで、ユーザのことを理解して話題提供したり、傾聴する対話を実現している。ユーザインタフェースは、一般的な PC 上で動作する Web ブラウザ用と、スマートフォン向けのものを準備している。ユーザのログイン、発言文/アンケート回答入力、対話や履歴表示などを行うフロントエンドは JavaScript、HTML、

CSS により構築し、チャット機能やストレス状態の解析を行うバックエンドは Python で主に構築した。また、本研究で開発したストレス検出システムの流れを図 3 に示す。



図 2. ストレス管理 AI チャットシステム(SMACS)の概要 (右：UI の例)

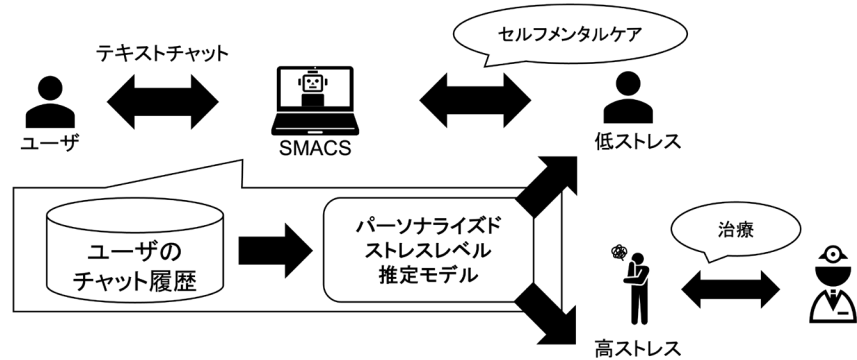


図 3. ストレス検出システムの流れ

SMACS のデータベースに蓄積された自己申告ストレスレベル（アンケート回答）と AI とのチャット履歴をもとにストレス検出モデルを構築する。このモデルによりストレス状態の自動検出を行い、低ストレス者には、SMACS の機能によってストレスの蓄積を防ぐためにセルフメンタルケアを促す。高ストレス者には、カウンセラーへの相談ルート確保を行う。カウンセラーは該当ユーザの情報を確認できるため、カウンセリング業務の効率化が期待できる。現段階では、高ストレス者の検出およびカウンセリング業務効率化機能は未実装である。テキストデータを利用したストレス推定の研究は存在するものの、実用化に向けて課題が残っている⁴⁾⁵⁾。本研究では、ストレス検出モデルはテキストチャット履歴を特徴量、ユーザが回答したストレス尺度をもとに算出したストレスレベルを正解データとして機械学習アルゴリズムを用いて構築する。ストレスの感じ方はユーザごとに個人差があるため、すべてのユーザに対応可能な汎用的なストレス検出モデルを構築することは困難であり、高い精度は望めない。しかし、ユーザごとに個人適応したストレス検出モデルを構築することで一定の性能が保てると考える。カウンセリング業務効率化のための分析データ共有機能は、カウンセラーの業務内容と判断基準を参考に今後、設計を進めていく予定である。

4. 評価実験と考察

データ収集は2024年8月から11月に実施し、徳島大学、香川高等専門学校、青森大学の3機関の学生（計22名）を対象に実施した。実験は以下の3つの手順で実施される。

本研究におけるデータ収集は、徳島大学の研究倫理委員会の承認を経て実施されており、被験者には、7日分のデータ入力ごとに謝金を支出した。

1. チャット前アンケートの実施、対話AIのタイプ選択
2. 対話AIとの10対話以上（システム・ユーザがそれぞれ5発話ずつ）のチャット
3. チャット後アンケートの実施

データ収集の結果、学生22名の参加者から計740件の有効回答が得られた。1人あたりの平均入力日数は約33.6日であった。表1に、システムを通して得られたアンケート結果の基本統計量を示す。各項目における平均値(mean)、標準偏差(std)、最小値(min)、第1四分位数(25%)、中央値(50%)、第3四分位数(75%)、最大値(max)を確認している。

表1. アンケートの基本統計量の一部

変数名	mean	std	min	25%	50%	75%	max
beforestress	2.31	1.12	1	1	2	3	5
beforeseintiment	0.23	1.41	-3	-1	0	1	3
beforevibrant	2.56	1.31	1	1	3	3	5
beforefun	2.65	1.34	1	1	3	4	5
beforefulfilled	2.74	1.37	1	1	3	4	5
beforedisturbed	1.59	0.9	1	1	1	2	5
beforeanxious	1.6	0.93	1	1	1	2	5
beforeagitated	1.49	0.85	1	1	1	2	5
beforecalm	2.66	1.05	1	2	3	3	5
beforeslow	2.64	1.07	1	2	3	3	5
beforepeaceful	2.65	1.08	1	2	3	3	5
afterstress	2.06	1.06	1	1	2	3	5
aftersentiment	0.47	1.3	-3	0	0	1	3
aftervibrant	2.68	1.31	1	1	3	3	5
after_fun	2.76	1.36	1	1	3	4	5
after_fulfilled	2.8	1.35	1	1	3	4	5
afterdisturbed	1.52	0.87	1	1	1	2	5
afteranxious	1.54	0.88	1	1	1	2	5
afteragitated	1.45	0.83	1	1	1	2	5
aftercalm	2.75	1.08	1	2	3	3	5
afterslow	2.74	1.08	1	2	3	3	5
afterpeaceful	2.77	1.09	1	2	3	3	5

次に、テキストチャット履歴の統計量を表2に示す。ユーザとシステムの発話を1日分のデータとし、McCabを用いて形態素解析を行った。その後、得られた形態素情報を基に、発話ご

との単語数および文字数を計算した。ユーザの発話の単語数や文字数に比べると、システム発話の単語数や文字数は約 5 倍以上になっている。

表 2. テキストチャット履歴における統計量

統計量	mean	min	25%	50%	75%	max
ユーザ単語数	84.86	7	39.75	68	113	407
ユーザ文字数	152.28	19	73.75	121	204	1116
システム単語数	491.66	74	324.75	440	609.5	1894
システム文字数	818.81	115	537.50	728	1019.5	4150

次に、うつ病や不安障害などの精神疾患をスクリーニングする質問票である K6⁶⁾の回答結果を集計した K6 スコアについて統計量を求めたものを表 3 に示す。K6 スコアは、5 点未満の場合は問題なし、5 点以上 10 点未満の場合なら要観察、10 点以上ならば要注意/要受診と判断できる。before_k6_score の平均値が 3.98 であるのに対し、after_k6_score は 2.47 と低下している。この平均値の低下は、テキストチャットによるストレス軽減を示唆している。

表 3. K6 スコアの統計量

統計量	mean	std	min	25%	50%	75%	max
before_k6_score	3.98	4.98	0	0	2	6	23
after_k6_score	2.47	3.62	-3	0	1	4	21

次に、ストレスの推定モデルを以下に示す 2 通りの条件に分けた実験により行った。対話前の K6 スコアの値をもとに 0 または 1 のラベルを付与し、このラベルを推定する。K6 スコア (before_k6_score) の閾値を 5 として、閾値未満は 0 (低ストレス)、閾値以上は 1 (高ストレス) として 1 対話ごとにラベルを付与した。

条件 1： 汎用型ストレス推定。すべてのユーザのデータを学習データに含めることでストレス推定モデルを学習する。

条件 2： 個人最適化ストレス推定。ユーザごとにストレス推定モデルを学習する。

1 対話におけるユーザ発話集合を Sentence-BERT⁷⁾を用いてベクトル化し、平均ベクトルを求めたものを特徴量とし、勾配ブースティング決定木に基づく機械学習アルゴリズム LightGBM⁸⁾を用いて 2 値分類モデル (低ストレス/高ストレスを分類) を構築した。データはクラス不均衡 (低ストレスが多数派クラス) であるため、クラスの重み付けおよびオーバーサンプリング手法である SMOTE⁹⁾、ランダムオーバーサンプリングを用いて学習データのクラスバランスを調整した。

表 4 に、LightGBM の学習において使用したハイパーパラメータの設定を示す。

表 4. LightGBM のハイパーパラメータ

パラメータ	設定値	説明
objective	binary	2 値分類タスクの指定
metric	binary_logloss	2 値分類用の対数損失による評価指標
boosting_type	gbdt	勾配ブースティング決定木の使用
num_leaves	31	1 つの決定木における最大葉ノード数
learning_rate	0.05	モデル更新時の学習率
feature_fraction	0.9	各イテレーションで使用する特徴量の割合
min_data_in_leaf	20	各葉ノードに必要な最小データ数
max_depth	-1	木の深さの制限なし

ストレス推定の評価指標には、適合率(Precision)、再現率 (Recall)、F 値 (F1-score) の 3 種類を用いる。Precision は、推定結果の精度を測る際に重要な指標であり、特に誤検出を避けるべきタスクでは重要となる。Recall は、特に見逃しを最小限に抑えるべきタスクで重要となる。F1-score は、適合率と再現率の調和平均を示す指標であり、両者のバランスを評価する際に用いられる。表 5、表 6 に条件 1 の汎用型ストレス推定モデル、条件 2 の個人最適化ストレス推定モデルにおける評価結果を示す (P:Precision, R:Recall, F1:F1-score)。個人最適化モデルにおいて、学習データをユーザごとに準備するためデータの量が非常に少なくなる。そのため、入力データの日数が少ないユーザ (14 日未満) は対象外とし、オーバーサンプリング法も、学習データ数がある程度必要な SMOTE ではなく、より単純なランダムオーバーサンプリングを用いた。

表 5. 汎用型モデルの評価結果

ラベル	クラスバランス調整なし			重み調整(class weight)			SMOTE		
	P	R	F1	P	R	F1	P	R	F1
0	0.68	0.98	0.80	0.68	0.90	0.78	0.74	0.88	0.80
1	<u>0.89</u>	0.29	0.44	0.69	0.34	0.46	0.73	<u>0.52</u>	<u>0.61</u>
Average	<u>0.77</u>	0.71	0.66	0.68	0.68	0.65	0.74	<u>0.74</u>	<u>0.73</u>

表 6. 個人最適化モデルの評価結果

ラベル	クラスバランス調整なし			重み調整(class weight)			Random Oversampling		
	P	R	F1	P	R	F1	P	R	F1
0	0.46	<u>0.54</u>	<u>0.50</u>	0.45	0.50	0.46	<u>0.53</u>	0.51	0.49
1	0.35	0.46	0.39	0.36	0.52	0.41	<u>0.46</u>	<u>0.59</u>	<u>0.50</u>
Average	0.68	<u>0.81</u>	<u>0.73</u>	0.64	0.75	0.68	<u>0.75</u>	0.75	<u>0.73</u>

以上の結果から、学習データがある一定の条件を満たした状態 (データの入力日数が 14 日以上ユーザ) の場合には、個人最適化モデルが汎用型モデルよりも高い再現率 (Recall) を達成できることがわかった。しかし、高ストレスの Recall は最高で 0.59 であり、現時点では実用的な性能に達していないことがわかる。

5. まとめ

本研究はストレスを抱える人々の支援を目的として開発したストレス管理 AI チャットシステムで収集したデータをもとに、個人適応可能なストレス検出モデルを構築することを目的とした。本報告では 2024 年 8 月から 11 月に実施した実験で収集したデータをもとに、各ユーザのチャット履歴とストレスレベルを測るアンケートの分析結果からストレスレベル推定モデルを構築し、収集したデータの妥当性を検証した。結果として、個人最適化型（パーソナライズド）ストレス推定モデルの平均性能は汎用型ストレス推定モデルの性能を上回る結果となった。本研究の途中経過については、発表文献¹⁰⁾¹¹⁾¹²⁾に詳細を記している。

また、本システムのチャット機能を利用することで、ストレスを軽減する効果がみられた。本研究で得られた結果から、以下のような課題が挙げられる。

収集データの不足および偏り

データ収集における課題として、システムの使用性に起因するデータ量の不足と偏りが挙げられる。システムの使いづらさにより、ユーザの継続的な利用が妨げられ、特に高ストレス状態のラベルデータの取得が困難となっている。さらに、高ストレス状態にある人々はそもそもシステムの利用自体を避ける傾向にある可能性も考えられ、このことがデータの偏りをより顕著にしている可能性がある。また、現状では学生を主な対象としてデータ収集を行っているため、より幅広い年齢層からのデータ収集を行い、システムの汎用性を検証する必要がある。

正解ラベルの信頼性

現在の K6 指標は主観的な自己報告に依存しているため、無意識のストレス状態を適切に検出するのが難しい問題がある。これを解決する手段として、定期的なフィードバックをユーザに提供し継続的な自己評価の精度を向上させることも有効である。

GPT を利用したシステムの再現性

プロンプトによる指示を与えても、期待する形式と異なる応答が返されることがあり、システムの安定性と信頼性に影響を与えている。期待される形式とは異なる応答が得られる問題を解決するため、モデルの応答の一貫性を検証するための自動評価システムを開発し、フィードバックに基づきモデルを継続的に改善するプロセスを導入することが望ましい。

時系列解析による特徴量の作成

時系列解析を活用した特徴量生成の有効性について、深層学習や機械学習の手法を用いたさらなる検証が求められる。具体的には、ストレス状態が変化する前兆となるパターンを識別するモデルを開発し、検証することで、効果的なストレス検出を実現することが期待される。

謝辞

本研究を遂行するにあたり、公益財団法人 天野工業技術研究所から多大なご支援を頂きました。ここに記して謝意を示します。共同研究者である労働者健康安全機構 木内敬

太氏、徳島大学大学院 康鑫 講師、香川高等専門学校 篠山学 准教授には様々なご助力をいただきました。感謝申し上げます。さらに、青森大学 藤澤日明 講師には、データ収集のための被験者募集にご尽力いただきました。ここに謝意を表明いたします。

参考文献

- 1) 長野県, LINE 株式会社. 長野県と LINE 株式会社による LINE を活用したいじめ等相談の中間報告資料, <https://d.line-scdn.net/stf/linecorp/ja/pr/NaganoPrefectureReportMaterial.pdf> (最終アクセス日 : 2025/6/23)
- 2) 高階光梨, 鈴木ひかり, 白塚龍太郎, 大橋佳奈, 宮下太陽, 横光健吾. 日本における抑うつ症状に対する心理学的支援を目的としたスマートフォン用アプリケーション・プログラムのレビュー, 認知行動療法研究, 2021, 47, 1, 1-10.
- 3) SELF 株式会社, <https://self.software/> (最終アクセス日 : 2025/6/23)
- 4) 岡本宗一郎, 砂山渡, 服部峻. ChatGPT を用いた SNS ユーザのストレスを含むコメントの抽出とストレス推定, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会(第 31 回) SIG-AM-31-03, p17-22, 2023, <https://must.c.u-tokyo.ac.jp/sigam/sigam31/sigam3103.pdf>
- 5) 深澤佑介. 人が生み出すビッグデータと機械学習によるメンタルヘルスの推定. 日本神経回路学会誌, 2022, 29, 2, 78-94, (DOI:10.3902/jnns.29.78)
- 6) Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.-L.T., Walters, E.E., & Zaslavsky, A. (2002). Short screening scales to monitor population prevalences and trends in nonspecific psychological distress. *Psychological Medicine*. 32(6), 959-976.
- 7) Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 3982–3992. (DOI:10.18653/v1/D19-1410)
- 8) Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 3149–3157.
- 9) Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)*. (DOI:10.1613/jair.953)
- 10) 森大樹, 松本和幸, 康鑫, 篠山学, 木内敬太. 社会的孤立の解消を目指したストレス管理 AI チャットシステムの開発, IT ヘルスケア誌 第 17 回年次学術大会抄録集, 19, 1, 100-105, 2024 年 8 月.
- 11) 森大樹, 松本和幸, 康鑫, 篠山学, 木内敬太. ストレス管理 AI チャットシステムによる個人適応可能なストレス検出モデルの構築, 第 23 回情報科学技術フォーラム(FIT2024)講演論文集, 3, 357-360, 2024 年 9 月.
- 12) Mori, D., Matsumoto, K., Kang, X., Sasayama, S. & Kiuchi, K. (2024). SMACS: Stress Management AI Chat System, In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024)*, 2, 167-174, Porto, Nov. 2024. (DOI: 10.5220/0012940300003838)