

ハッカソン 参考資料

UTTC 7期生(AIコース)

大宮 拓朝

UTokyo
Tech
Club

目次

1. はじめに

2. ユーザーの解約予測AIの実装

3. Gmailを用いたAIアプリの実装

3.1. アイデア・市場分析

3.2. バックエンドの工夫点

3.3. フロントエンドの工夫点

3.4. デモ動画

4. おわりに

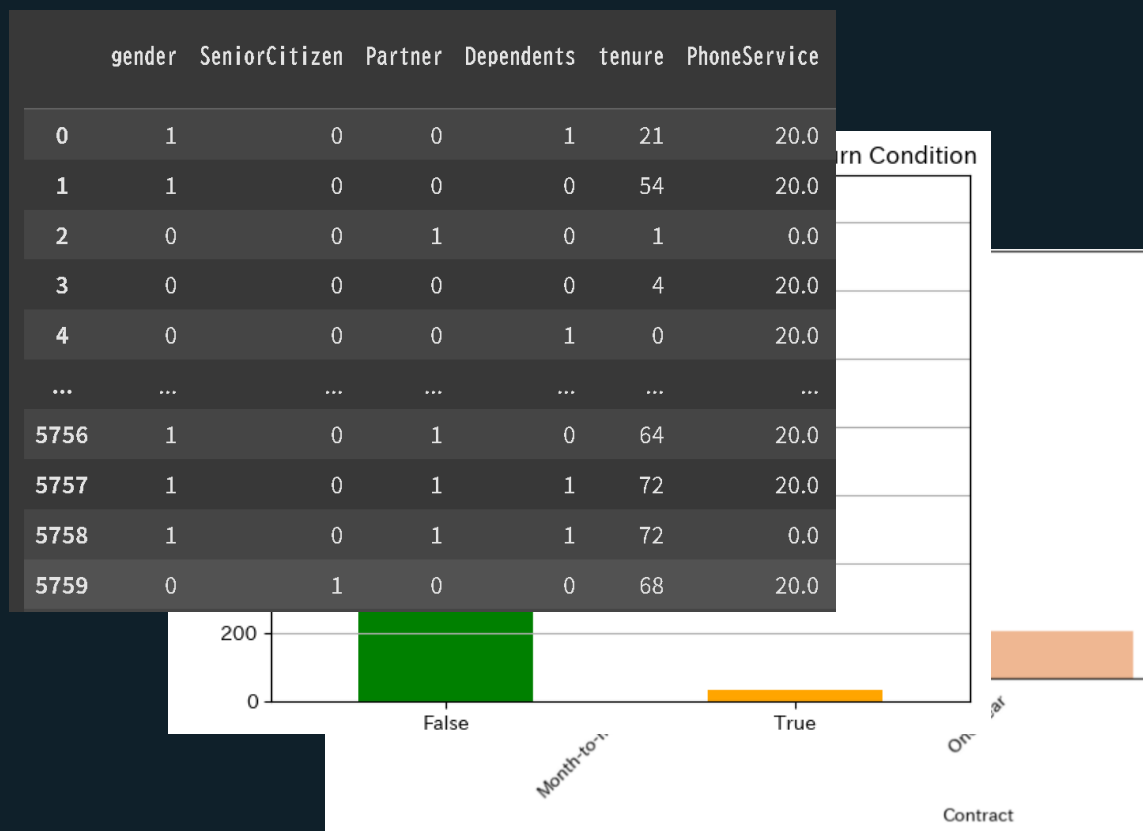
1. はじめに

1. はじめに

課題① 解約予測AI

最終CV F1スコア : 0.63444

テスト F1スコア : ??



課題② Gmail AIアプリ

高齢者向けメールアプリ「FUMI」
(3大機能: 迷惑mail/見守り/AIアシスト)



2. ユーザーの解約予測AIの実装

2. ユーザーの解約予測AIの実装

主に行った工夫は以下の5点です。

- ①ヒューリスティックな手法を用いた**新変数の模索**
- ②**重回帰分析**を用いたカテゴリデータのラベリング
- ③マハラノビス距離/One-Class SVM Scoreを用いた**異常検知**
- ④機械学習モデル6つを使った**アンサンブリング**
- ⑤閾値やアンサンブリング時の重みづけ平均の**最適化**

①ヒューリスティックな手法を用いた**新変数の模索**

EDAでの解析により、F1スコアが上がった変数が見つかった。
以下は、その中でも特にF1スコア向上に繋がった新変数である。

- ・ 契約期間が12カ月未満に関してフラグ立てを行うと良い
⇒ **短期間でサービスに不満足で解約する**人の特徴を捉えた
- ・ 契約形式(1カ月更新/1年更新/2年更新)と累計契約期間から、**次回契約更新月までの月数**を計算すると良い
⇒ 残り月数が0に近いほど解約しやすい傾向を捉えた
- ・ 「月額料金×契約期間－累計支払料金」を計算すると良い
⇒ これがほぼ0の顧客は**サービス内容の途中変更がなく、**
0から離れている場合は**サービス内容を変更**した可能性あり

②重回帰分析を用いたカテゴリデータのラベリング

- 一般的なカテゴリデータのエンコーディング手法として、LabelEncoding等があるが、精度が伸びず。
- EDAを進めたところ、カテゴリデータのうち9変数は「MonthlyCharges」に依存することが分かった。
- 以下のように重回帰分析を行ったところ、決定係数0.99超月額総額を各サービスの単価に分解できたと考えられる。こちらでF1スコア向上

OneHotEncoding後に重回帰分析

customerID

gender

PhoneService

MultipleLines

InternetService

OnlineSecurity

OnlineBackup

DeviceProtection

TechSupport

StreamingTV

StreamingMovies

MonthlyCharges

TotalCharges

Churn

0

4223-BKEOR

Female

Yes

No

DSL

Yes

No

Yes

No

No

Yes

64.85

1336.8

No

1

6035-RIIOM

Female

Yes

Yes

Fiber optic

No

Yes

No

No

Yes

Yes

97.20

5129.45

No

2

3797-VTIDR

Male

No

No phone service

DSL

No

No

No

No

No

No

23.45

23.45

Yes

重回帰分析の結果で得た回帰係数でencoding

③ マハラノビス距離/One-Class SVM Scoreを用いた異常検知

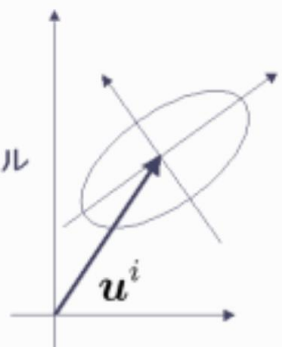
- 説明変数のみ(エンコーディング済)のデータの分散共分散行列を計算し、**マハラノビス距離**を計算した
- マハラノビス距離は多変数の**外れ値**を検出でき、F1スコア向上に貢献
- ~~One-Class SVM Score(SVMを教師なしの1クラス分類に応用した手法)で、異常なし(1)と異常あり(-1)に分類できる(2025.6.21変更)~~

共分散行列

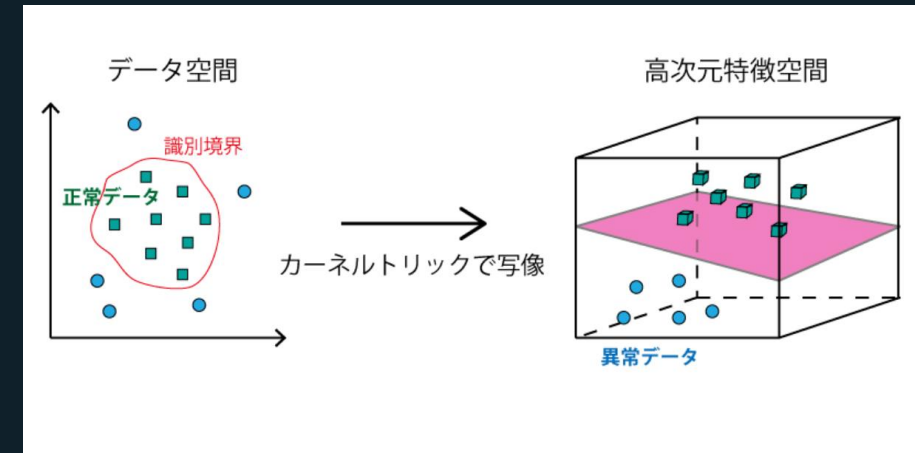
$$D_M(x, u^i) = (x - u^i)^T \Sigma_i^{-1} (x - u^i)$$
$$= \sum_{j=1}^n \frac{((x - u^i)^t \phi_j)^2}{\lambda_j}$$

固有値 固有ベクトル

x : 未知入力ベクトル
 u^i : 平均ベクトル
 n : 次元数



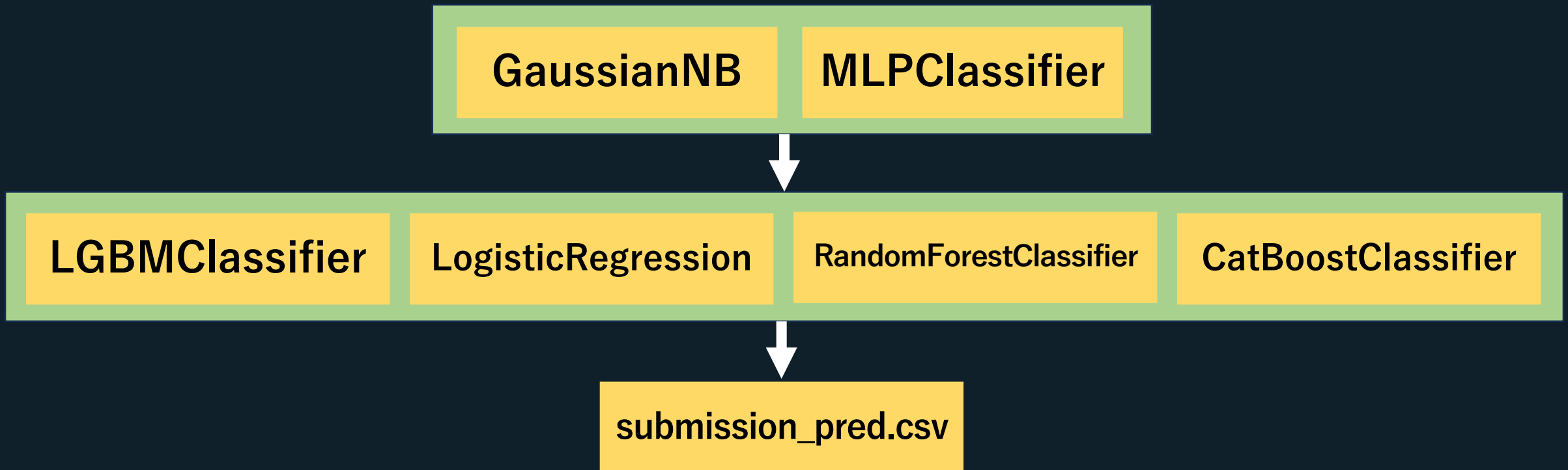
<http://www.it.is.tohoku.ac.jp/~kato/patman/sld031.htm> より引用



<https://hktech.hatenablog.com/entry/2018/10/11/235312> より引用

④機械学習モデル6つを使ったアンサンブリング

- ・ 下記の様なアンサンブル方法を使用した。
 - ① GaussianNB/MLPClassifierの予測結果を**新変数で追加**
 - ② ①で追加した変数も含めてLGBM等の**4モデルで予測**
 - ③ 4つのモデルの**重みづけ平均**で最終的な予測値を取得



⑤ 閾値やアンサンブリング時の重みづけ平均の最適化

- 最終アンサンブル時に4つのモデルのアンサンブルを行う際、**trainデータを最もよく分類できる比率**で重みを決定する

LGBMClassifier

LogisticRegression

RandomForestClassifier

CatBoostClassifier


[0.07652765 0.69651854 0.12376249 0.10319131]

- 重みづけ平均で得た解約可能性(%)について、何%を境に解約/非解約の判定をする**最適な閾値を決定する**
※0.50で行う場合よりもF1スコアが0.005ほど微増する

3. Gmailを用いたAIアプリの実装

3.1. アイデア・市場分析

- ・「Gmailを用いたAIアプリ」が要件だった。
⇒他のSNSにない“メール特有の何か”を捉えて差別化したい

- 
- ・カジュアル
 - ・若年層/中年層
 - ・オープンな感じ
 - ・リアルタイム性

- ・フォーマル
- ・中年層/高齢者層
- ・クローズドな感じ
- ・持続的な記録



- ・主なターゲット層を高齢者としたメールサービスを考える

3.1. アイデア・市場分析

- ・ 高齢者のPCでのEメール利用は一定数見込める。
- ・ 詐欺対策をどのように行うかがネックになりそう。
- ・ メールを通して高齢者の生成AI活用の足掛かりにしたい。

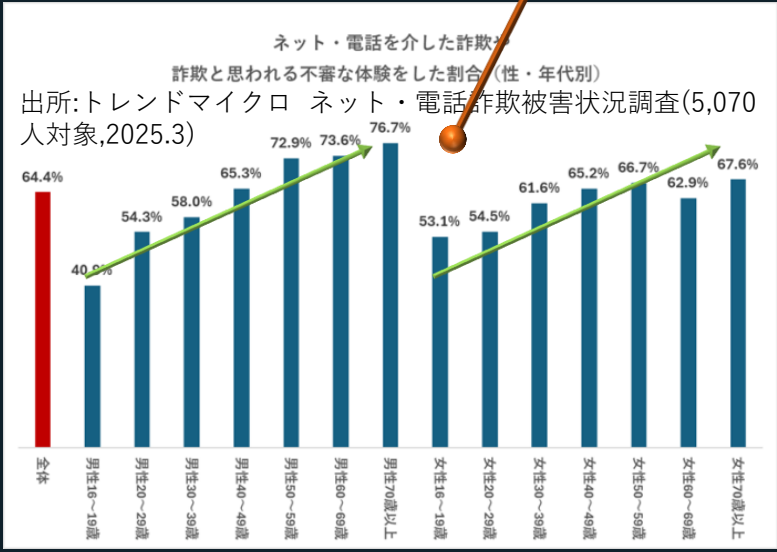
＜市場分析①＞
高齢者程PCでのメール利用率 ↑

表3. パソコンでのサービス利用率（年代別）
[調査対象：全国・15～79歳のパソコン所有者・n=3,844]

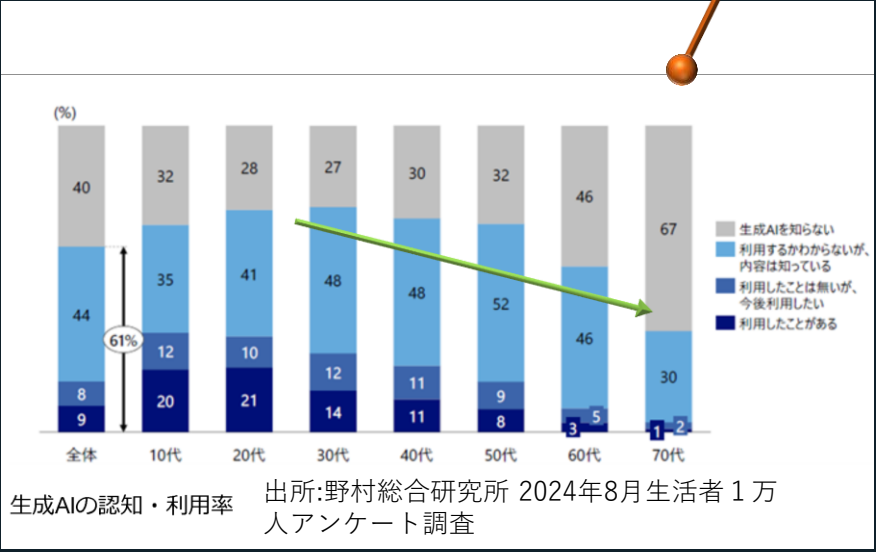
	全体(%)	全体からの差分(pt)		
		10代～30代 (n=1,109)	40代～50代 (n=1,228)	60代～70代 (n=1,507)
電子メール	73	-17	2	11
情報検索・ニュースの閲覧	72	-14	4	7
ネットショッピング・フリマ	60	-19	4	10
動画や音楽	43	6	1	-5
地図・乗換・ナビ	34	-16	-1	12
ソーシャルメディア	28	2	2	-3
資格・勉強・習い事関連	15	9	-1	-6
ゲーム	15	8	-2	-4
LINEなどのコミュニケーションツール	14	5	-3	-1
健康関連サービス	11	-4	-1	4

出所:NTTドコモ モバイル社会研究所 2022年スマホ利用者行動調査

＜市場分析②＞
高齢者程ネットで詐欺に遭い易い

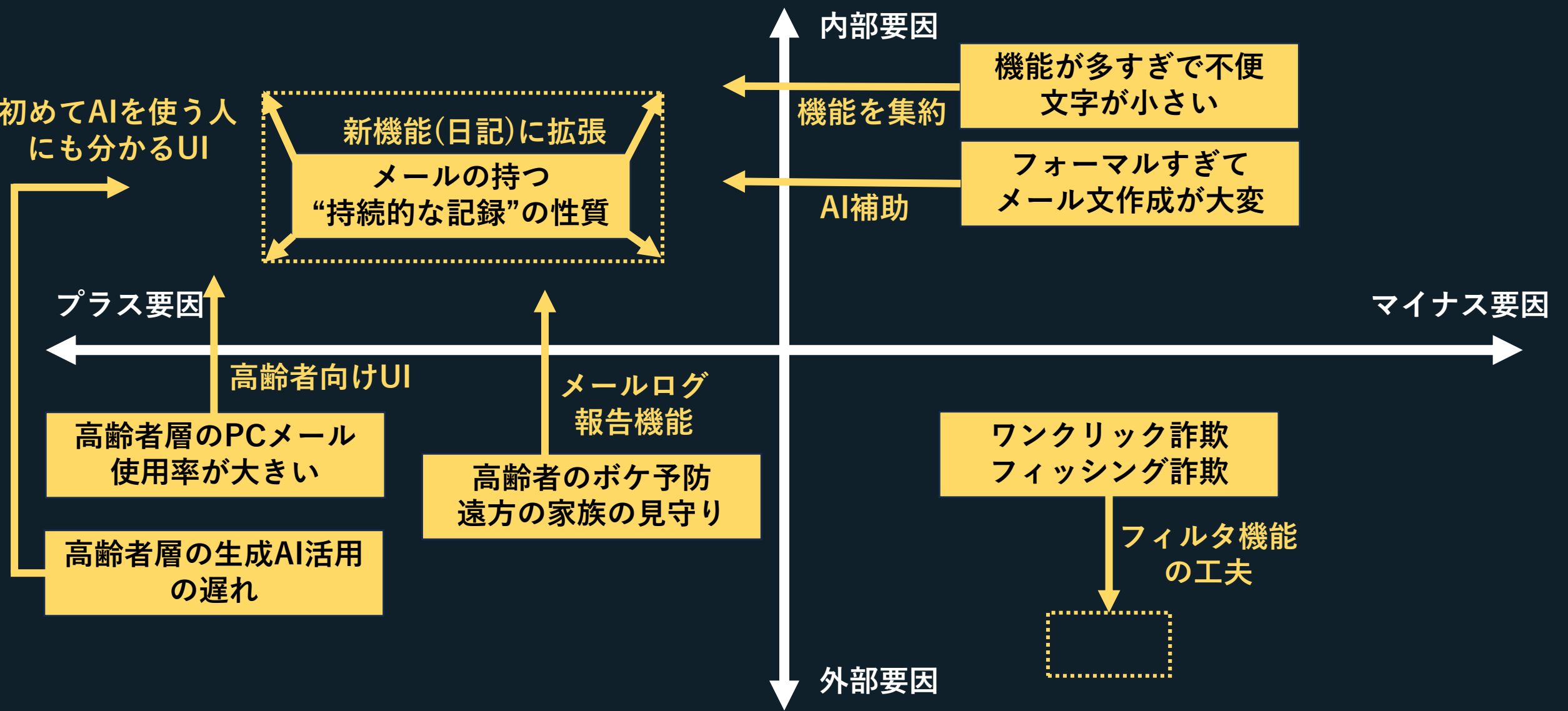


＜市場分析③＞
高齢者程生成AI活用が遅れている



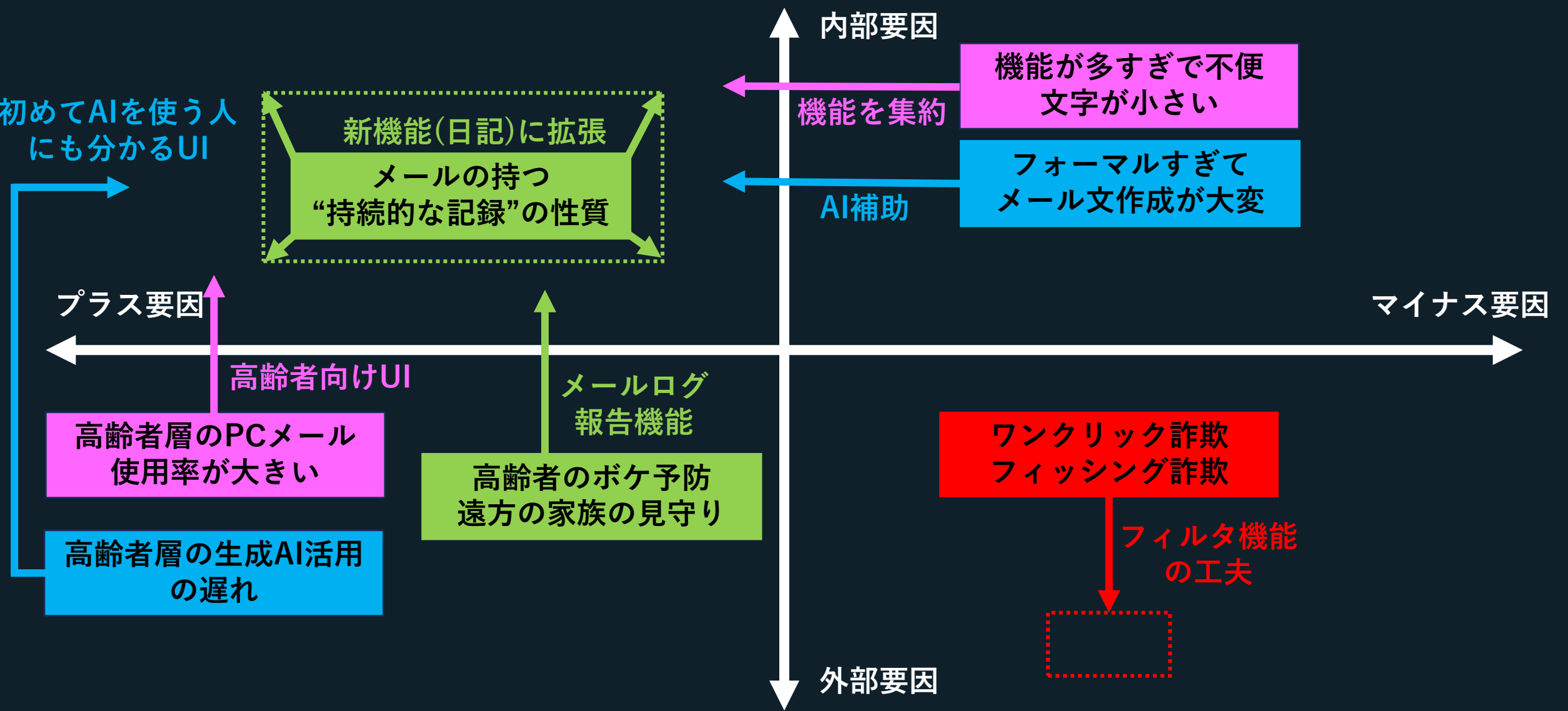
3.1. アイデア・市場分析

ターゲット層を高齢者とした時のEmailアプリ



3.1. アイデア・市場分析

ターゲット層を高齢者とした時のEmailアプリ



3.1. アイデア・市場分析



ホーム画面です

Gmailと連携

最新のメールを取得

<Point①>
高齢者フレンドリーなUI



新規メール



ゴミ箱



設定

From: "出前館" <support@demaie-can.com>

メールアドレスの確認/Confirmation of e-mail address

メールアドレス変更を受け付けました。本人確認の後、メールアドレスを変更します。以下の…

From: e-slip administrator <fk8wt279@kyuyo3.keitai-bin.jp>

<Point②>
受信メールの危険度を可視化

デモ用

【重要なお知らせ】お客様のアカウントは不正利用の可能性があります お客様のアカウントに…

From: Takuto OMIYA <omiya-takuto8128@g.ecc.u-tokyo.ac.jp>

△ 危険度 95%

戻る

デモ動画用

送信元: Takuto OMIYA <omiya-takuto8128@g.ecc.u-tokyo.ac.jp>

日時: 2025年6月20日金曜日 10:26

<Point③>
分かりやすいAI機能

本文:

こんにちは。先日の講演会でお世話になりました大宮です。6/22(日)16:00から渋谷でお話できないでしょうか。

返信

転送

削除

今日の体調

良好

睡眠時間 (時間)

8

<Point④>
日記機能で家族の見守りが可能

日記を登録

今日は犬の散歩をしたが、近所の人にも話しかけられて楽しかった。

日記を登録

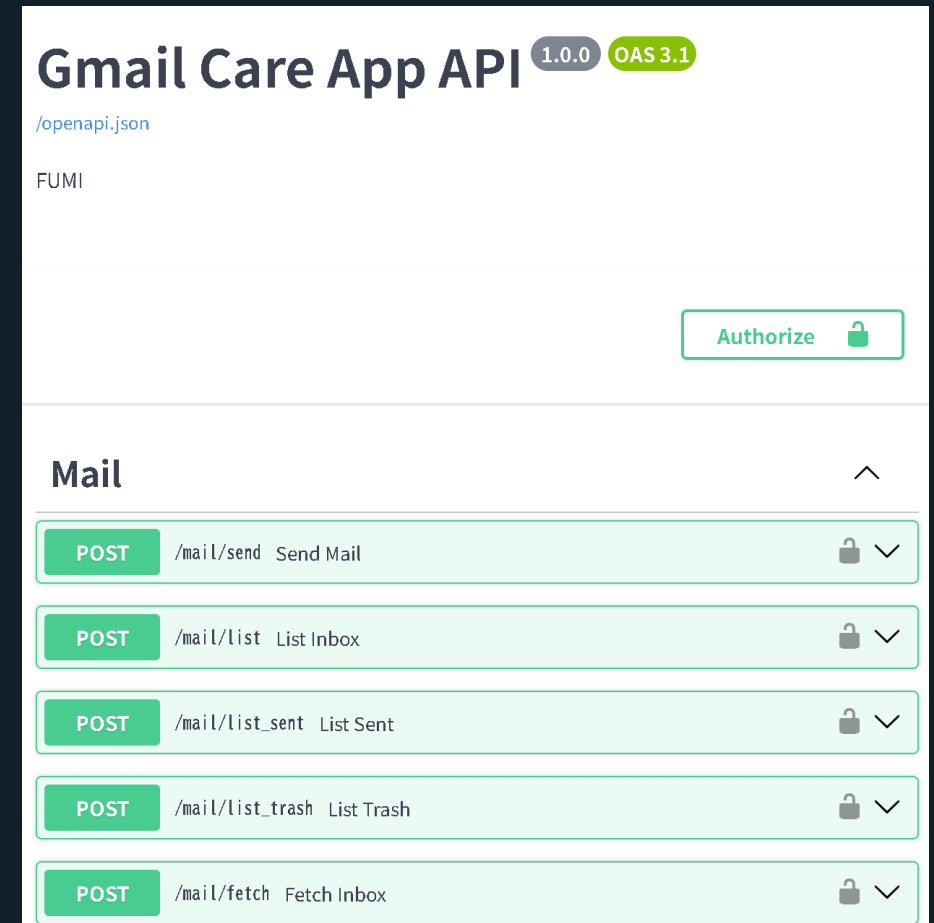
3.2. バックエンドの工夫点

ログイン画面等で求められる**パスワード**は**hash化**することで秘匿性を高めている



A login form for an application named "FUMI". At the top is a logo featuring a bird and the word "FUMI". Below the logo is a text input field containing the email address "t.omiya496@gmail.com". Underneath the email field is a password input field represented by a series of dots. At the bottom of the form is a large blue button with the text "ログイン" (Login) and a hand cursor icon pointing at it.

メールの送受信等が行われるルートはプライバシーの観点から**トークン認証**を行うようにした



The interface for the "Gmail Care App API". At the top, it says "Gmail Care App API" with version "1.0.0" and "OAS 3.1" badges. Below this is a link to "/openapi.json" and the text "FUMI". On the right side, there is a green "Authorize" button with a lock icon. Below the header is a section titled "Mail" with a collapse icon. This section contains a list of API endpoints, each with a green "POST" button, the endpoint path, a description, a lock icon, and a dropdown arrow.

Method	Endpoint	Description	Auth	More
POST	/mail/send	Send Mail	Lock	Dropdown
POST	/mail/list	List Inbox	Lock	Dropdown
POST	/mail/list_sent	List Sent	Lock	Dropdown
POST	/mail/list_trash	List Trash	Lock	Dropdown
POST	/mail/fetch	Fetch Inbox	Lock	Dropdown

3.2. バックエンドの工夫点

従来のように**2値分類**で迷惑メールを振り分けると、フォルダに振り分けられなかったスパムメールに対して無力

Gmailの迷惑メールフィルタを通過したメールの中でも、怪しい内容のものを**定量的な数値で表現**し赤くすることで相対的な危険度が分かるようにした

迷惑メール



From: "出前館" <support@demae-can.com>

メールアドレスの確認/Confirmation of e-mail address

メールアドレス変更を受け付けました。本人確認の後、メールアドレスを変更します。以下の…

From: e-slip administrator <fk8wt279@kyuyo3.keitai-bin.jp>

アプリテスト用

【サービス改善に関するアンケートのお願い】 平素より〇〇サービスをご利用いただき、誠に…

From: Takuto OMIYA <omiya-takuto8128@g.ecc.u-tokyo.ac.jp>

デモ用

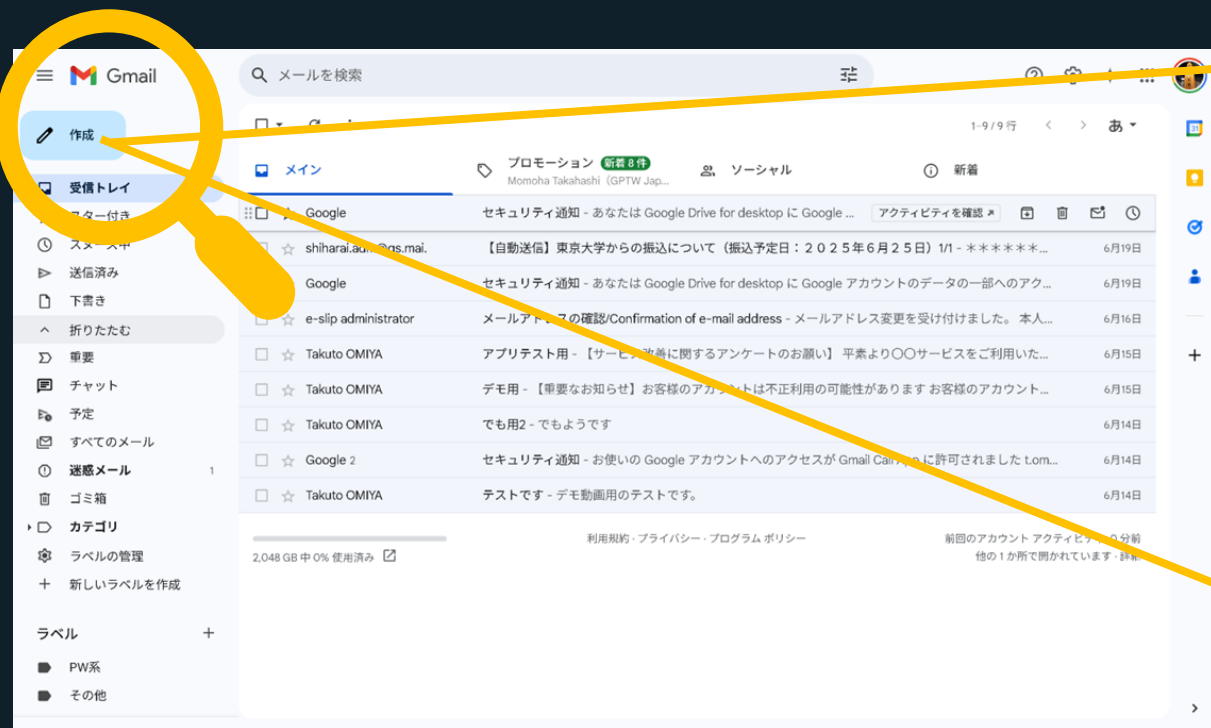
【重要なお知らせ】お客様のアカウントは不正利用の可能性があります お客様のアカウントに…

From: Takuto OMIYA <omiya-takuto8128@g.ecc.u-tokyo.ac.jp>

△ 危険度 95%

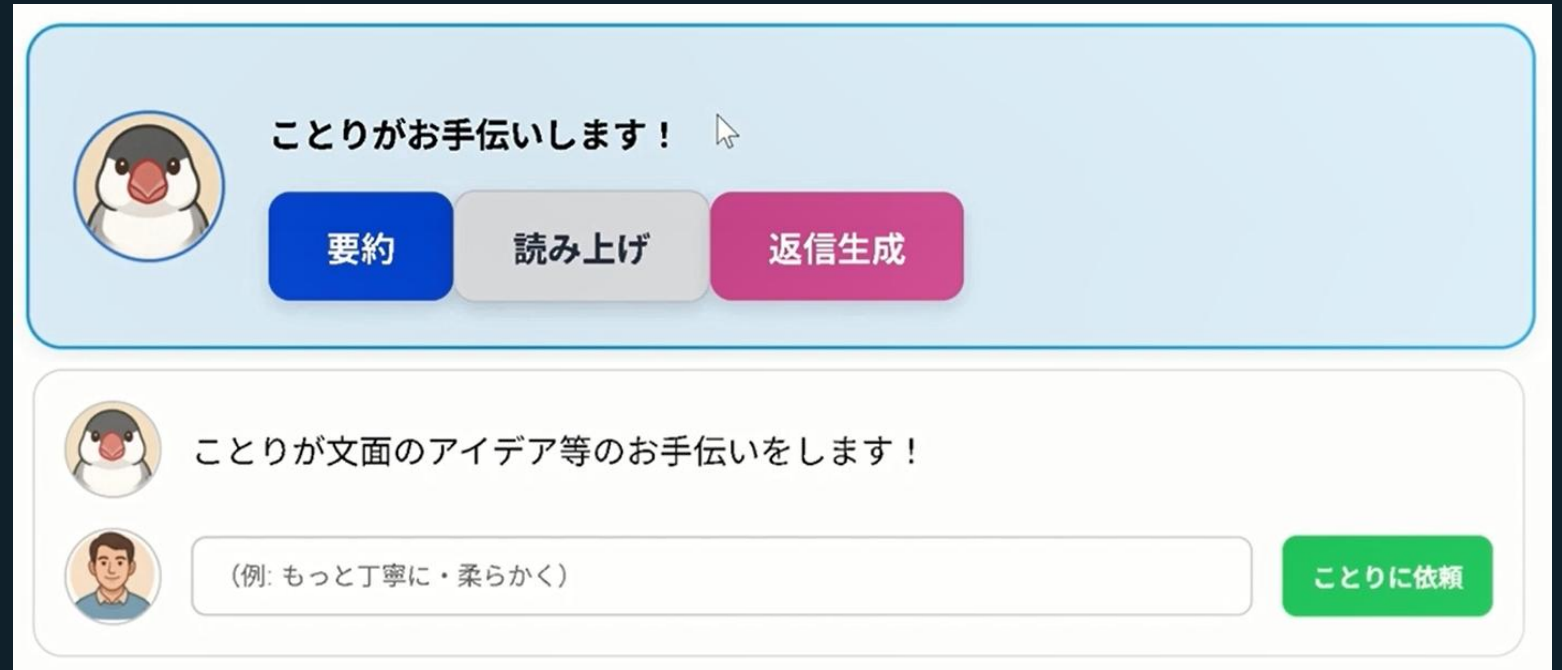
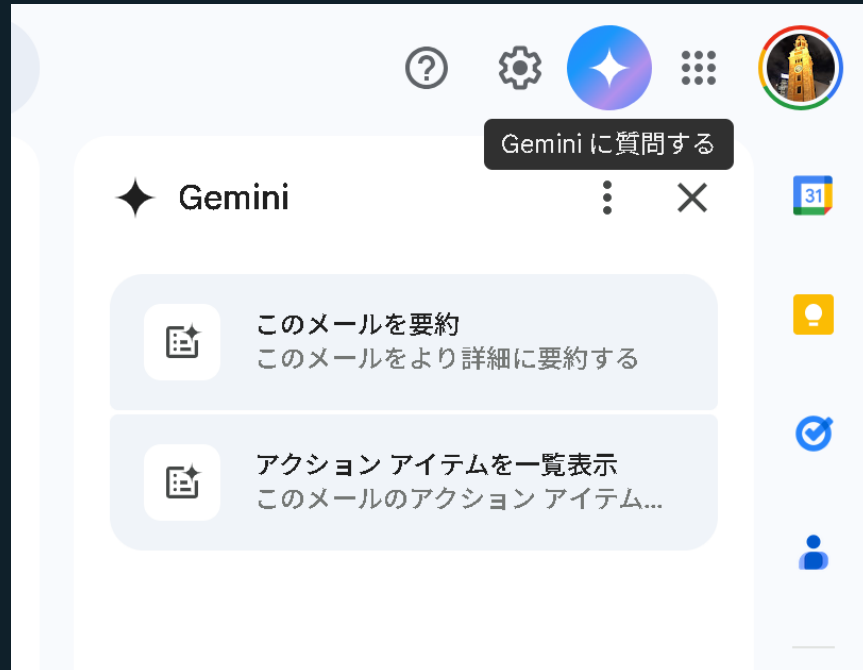
3.3. フロントエンドの工夫点

- ・ **「1画面1機能」** という考え方を重視。1画面毎のボタン数を少なくして、高齢者にわかりやすい設計・文字サイズを意識。
- ・ 高齢者のうち色の識別が付きにくい方向けに、**ホバーの動きを大きめ**にして、ボタンの使用感を良くした。



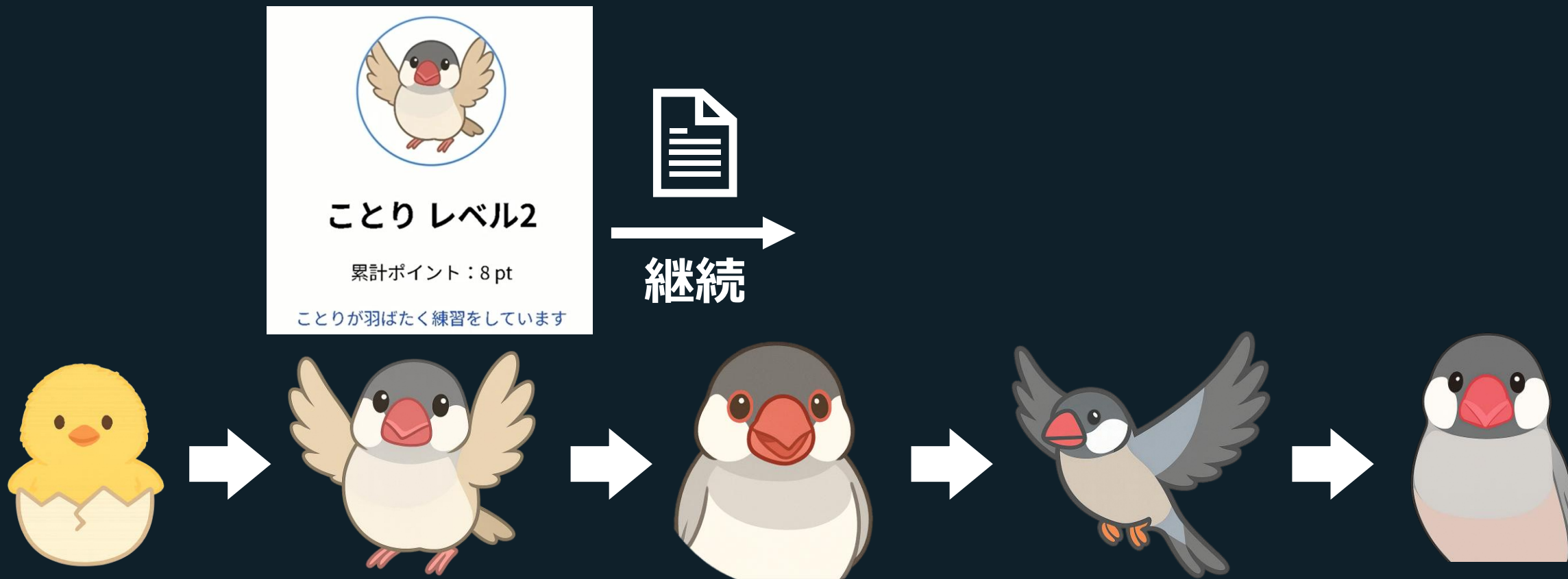
3.3. フロントエンドの工夫点

- ・従来の**無機質なAI**だと、“ロボット”感が出てしまう。
⇒AIが絡む処理は全て「ことり」というキャラクターを出す
ことで高齢者に馴染みのないAIに**親近感が湧く**ようにした
- ・AIに指示をテキスト入力する画面も「LINE」の様なUI設計
にすることで「指示を与える」操作を直感で分かるようにした



3.3. フロントエンドの工夫点

- ・従来の高齢者の見守り機能(GPS・カメラ設置等)は**“監視”的な側面**があり、導入に消極的な高齢者も多いらしい。
- ・**育成ゲーム的要素**を取り入れ、日記を継続することでキャラクターが成長する設計にし、見守り機能のマイナス面を軽減



3.4. デモ動画

ディレクトリ内にあるdemo.mp4をご覧ください

4. おわりに

4. おわりに

- ・ 3カ月間のカリキュラム/ハッカソンを通して、特にAIを実際のアプリに組み込む際の実装の全体感(フロントエンド・バックエンドの繋がり等)を学べたのが貴重でした。
- ・ 最終課題の内解約予測のAIに関しては、元々経験のあったKaggle形式のコンペの知識を活かし、0.4台だったF1スコアを0.63まで上げられたのが収穫でした。
- ・ Gmailを用いたAIアプリの開発に関しては、完全に初心者からのスタートだったため、ハッカソン期間中何度もAIコースの教材を見返し(フロントエンドは特にWebコースの教材も参考になりました)、オリジナリティのある作品ができました。ありがとうございました！