

株式会社Preferred Elements

研究開発

計画

- ①100B モデルの事前学習・追加学習
- ②100B モデルの指示学習
- ③画像モーダル向けの事前学習・追加学習
- ④音声モーダル向け追加学習
- ⑤1Tモデルの事前学習の検証

成果

- ①**事前学習完了。5/13時点でJaster 4shotで0.71を記録(*スライド4で説明)**
- ②SFT Full finetuningの検証、データセット生成、社内GPUクラスタを用いたマルチノードでの動作確認を実施
- ③LAIONデータセット公開中止による目標の修正、全学習ステージの大規模分散学習開始及び性能評価、引き続きデータセットの準備及び前処理
- ④Whisper-largeと100Bモデルの学習パイプライン構築。特にスナップショット機構や評価スクリプトの開発
- ⑤**MoEの各種アプローチによる学習手法の性能検証、1Tパラメータモデルの分散学習基盤の開発**

社会実装に向けた取り組み

成果物の公開



ソースコード

- ・ ファインチューニング用



モデル

- ・ 事前学習済100Bモデルウェイト



開発ノウハウ

- ・ マルチモーダル化及び1Tモデル学習ノウハウの共有

事業化・実用化

- ・ 基盤モデルのAPI 提供
- ・ 基盤モデルのライセンス提供
- ・ 基盤モデル及び APIの改良に伴うエンジニアリング・コンサルティング

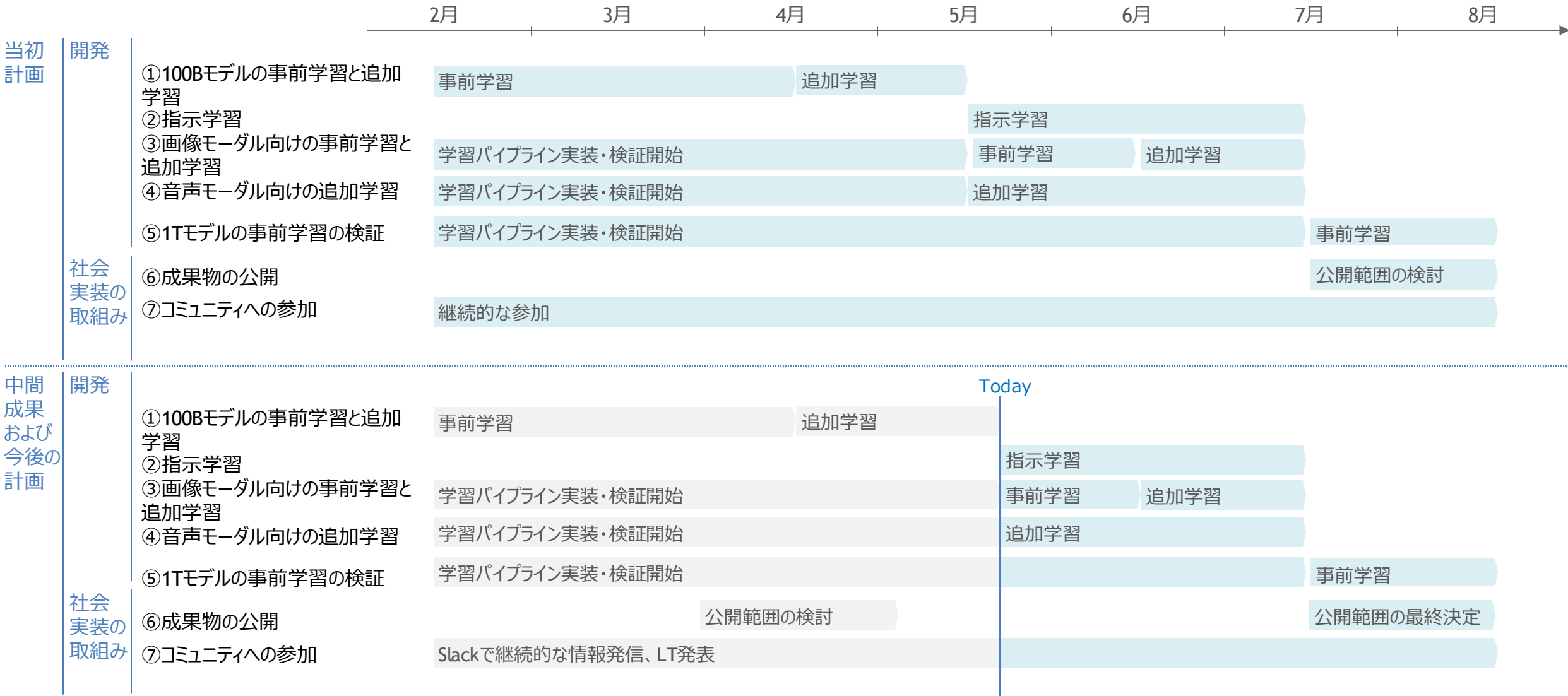
- ・ 公開範囲の検証
- ・ 開発期間終了後に、モデル自体は重みを含めて公開した上で、調査研究や学術用途での利用は無償、商用利用は有償とするライセンス設定予定
- ・ GENIAC SlackチャンネルやLTなどコミュニティで情報発信し、先行事業者として得た経験やtipsを他事業者にも共有
- ・ 有望顧客候補の選定、モデル提供に向けた計画の検討
- ・ GENIACマッチングイベントへの参加、顧客接点の拡大



今後の展望: 100Bモデル：指示学習を7月中旬まで約2か月実施し、**指示学習終了時は日本語ベンチマークで最高性能を狙う**
 1Tモデル：7月中旬から8月15日まで約1か月実施し、**MoE手法の確立、実装、最適化及び性能の確認を行う**

スケジュール

学習開始当初のGCPメンテナンスや不調による中断・復帰、GW期間中のスナップショットファイル破損はあったもののおおむね順調に進捗、**残りの期間も検証用計算器での事前検証・準備を周到に行い計画通り終了させる予定**



研究開発

事業 項目①

中間時点までの計画

- ・ 事前学習
- ・ 高品質日本語データによる追加学習

中間時点での成果

- ・ 事前学習完了。**5/13時点での暫定的な性能評価ではJaster 4shotで0.71を記録し、日本語性能としてはGPT-4(0.77)に迫る性能**
- ・ 高品質日本語データによる追加学習は5/24に終了を見込む
- ・ 事前学習の実効効率は**540TFLOPs/s/GPU**で40%目標（400TF/s/GPU）を達成

課題・乗り越え方

- ・ 自己負担で計算資源を調達し、助成期間開始前にGCP環境での学習に適用するための準備を進めていた
- ・ GCPメンテナンスや不調による中断・復帰が断続的に発生



今後の展望: 後続の事業項目②指示学習にて更なる性能向上

事業 項目②

中間時点までの計画

- ・ 事業項目①の後に実施予定のため、中間地点では実施項目なし

中間時点での成果

- ・ SFTに向けたQLoRA動作確認、データセット準備

課題・乗り越え方

- ・ -



今後の展望: 指示学習終了時に日本語ベンチマークで最高性能を狙う

事業項目①モデル性能（5/13時点）

- 事前学習途中+チューニング後の性能評価を実施
- 当社100Bモデル(PLAMO-100B)は、Jaster 4shotで0.71を記録しGPT-4(0.77)に迫る性能**
- 高品質日本語データによる追加学習は5/24に終了予定
- 追加学習と事業項目②指示学習によって更なる性能向上を予定している

	Jaster 4shotスコア
PLAMO-100B(5/13 暫定版)	0.71

*参考値（他モデルのGENIAC内でのスコア）	
Swallow 70B	0.68
StableBeluga2 (70B)	0.67
GPT-4	0.77

研究開発

事業
項目③

中間時点までの計画

- 画像モーダル向け学習パイプライン実装・検証開始

中間時点での成果

- 全ての学習ステージを含めた、大規模マルチクラスタ環境における学習の事前確認検証
- 物体検出モデル開発
- 日本語対応を含めたデータセット構築

課題・乗り越え方

- LAIONデータセットの公開停止による目標の調整



今後の展望: 全学習ステージの大規模分散学習開始及び性能評価、引き続きデータセットの準備及び前処理事前学習および追加学習の実施

事業
項目④

中間時点までの計画

- 音声モーダル向け学習パイプライン実装・検証開始

中間時点での成果

- Whisper-largeと100Bモデルの学習パイプライン構築
- 事前情報付き音声データセットの準備

課題・乗り越え方

- なし



今後の展望: Whisper-large+100B+QLoRAの学習の実施

研究開発・社会実装

事業 項目⑤

中間時点でのまでの計画

- 学習パイプライン実装・検証開始

中間時点での成果

- MoEの各種アプローチによる学習手法の性能検証
- 1Tパラメータモデルの分散学習基盤の開発

課題・乗り越え方

- ー



今後の展望: MoE手法の確立、実装、最適化及び性能の確認

ソース コード

中間時点でのまでの計画

- ー

中間時点での成果

- 公開範囲の検討
- 開発期間終了後に、モデル自体は重みを含めて公開した上で、調査研究や学術用途での利用は可能、商用利用は不可（有償）とするライセンスを設定する予定

課題・乗り越え方

- ー



- 今後の展望: 安全性や経済合理性の検証を行い、最終的な公開範囲を決定

社会実装

モデル	中間時点でのまでの計画	中間時点での成果	課題・乗り越え方
	－	- 公開範囲の検討 - 開発期間終了後に、モデル自体は重みを含めて公開した上で、調査研究や学術用途での利用は無償、商用利用は有償とするライセンスを設定する予定	－



今後の展望: 安全性や経済合理性の検証を行い、最終的な公開範囲を決定

開発 ノウハウ	中間時点でのまでの計画	中間時点での成果	課題・乗り越え方
	－	- GENIAC SlackチャンネルやLTなどコミュニティで情報発信し、先行事業者として得た経験やtipsを他事業者にも共有 3月LTでは、助成期間開始前にGCP環境での学習に適用するために行った準備についてスピーチ	－



今後の展望: GENIAC SlackチャンネルやLT会などイベント通じた情報発信の継続

実用化・事業化

実用化・
事業化

中間時点でのまでの計画

- 顧客候補へのコンタクト、実装に向けた交渉
- GENIACマッチングイベント参加

中間時点での成果

- 有望顧客候補の選定
- 顧客候補との具体的な案件の要件定義、プロジェクト化に向けた初期検討を実施中

課題・乗り越え方

- 他顧客、他案件の模索・拡大



今後の展望: マッチングイベントなど通じた顧客候補との接点の増加、案件パイプラインの拡大、モデル提供実現