

4th Grand Canvas: AI品質の未来を共に描く
～品質マネジメントネットワーキングシンポジウム～
@AP東京八重洲

Preferred Networksにおける大規模言語モデル 開発と活用での AI安全性の考え方

2025年3月4日(火)
株式会社 Preferred Networks

大野健太 (ono@preferred.jp)

目次

Agenda

- 1 会社概要
- 2 PFNのAIガバナンス体制
- 3 過去の機械学習サービスの品質管理事例
- 4 まとめ

Preferred Networks (PFN) 会社概要

ミッション：現実世界を計算可能にする

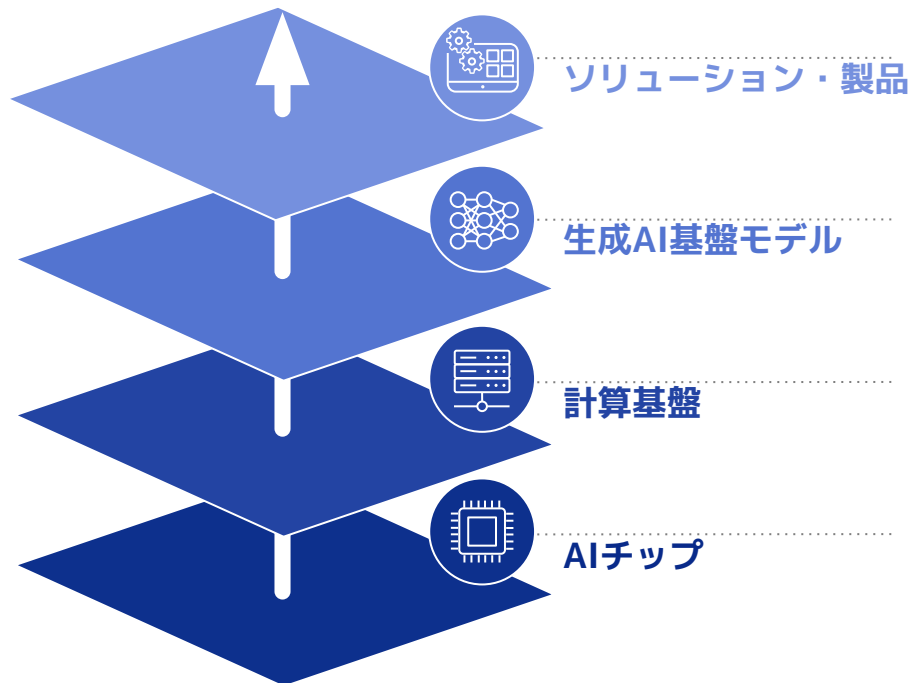
設立	2014年3月26日
本社	東京都千代田区
代表取締役	西川徹（最高経営責任者）、岡野原大輔（最高研究責任者）
従業員数	約350名（2024年9月）
事業内容	AIチップ、計算基盤、生成AI・基盤モデルなどのAI関連技術を活用したソリューション・製品の開発・販売および研究開発
主要子会社	株式会社Preferred Computational Chemistry（2021年6月） 株式会社Preferred Robotics（2021年11月） 株式会社Preferred Elements（2023年11月） 株式会社Preferred Computing Infrastructure（2025年1月）
出資企業	SBIグループ ENEOSイノベーションパートナーズ合同会社 積水ハウス投資事業有限責任組合 中外製薬株式会社 TEL Venture Capital, Inc. トヨタ自動車株式会社 株式会社日本政策投資銀行 日本電信電話株式会社 株式会社博報堂DYホールディングス 株式会社日立製作所 ファナック株式会社 株式会社みずほ銀行 三井物産株式会社 三菱商事株式会社 株式会社ワコム



<https://www.preferred.jp>

PFNの事業: AI技術のバリューチェーンを垂直統合

PFNは、チップ、計算基盤、生成AI基盤モデル、ソリューション・製品まで、AI技術のバリューチェーンを垂直統合し、ソフトウェアとハードウェアを高度に融合することで、競争力の高い技術の開発および産業応用を進めています。



様々な産業・消費者向けのソリューション・製品

PreferredAI

MiseMise

MATLANTIS

PFN 3D Scan

kachaka

Preferred Networks
Visual Inspection

rypko

PLaMo

PLaMo Prime (国産LLM)
PLaMo Lite (エッジ向けSLM)

PFP

物質のエネルギー計算モデル



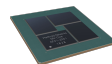
GPUクラスタ



MN-3
(MN-Core™
クラスタ)

PFCP

MN-Core™ 2を
計算資源とした
クラウドサービス



MN-Core™



MN-Core™ 2



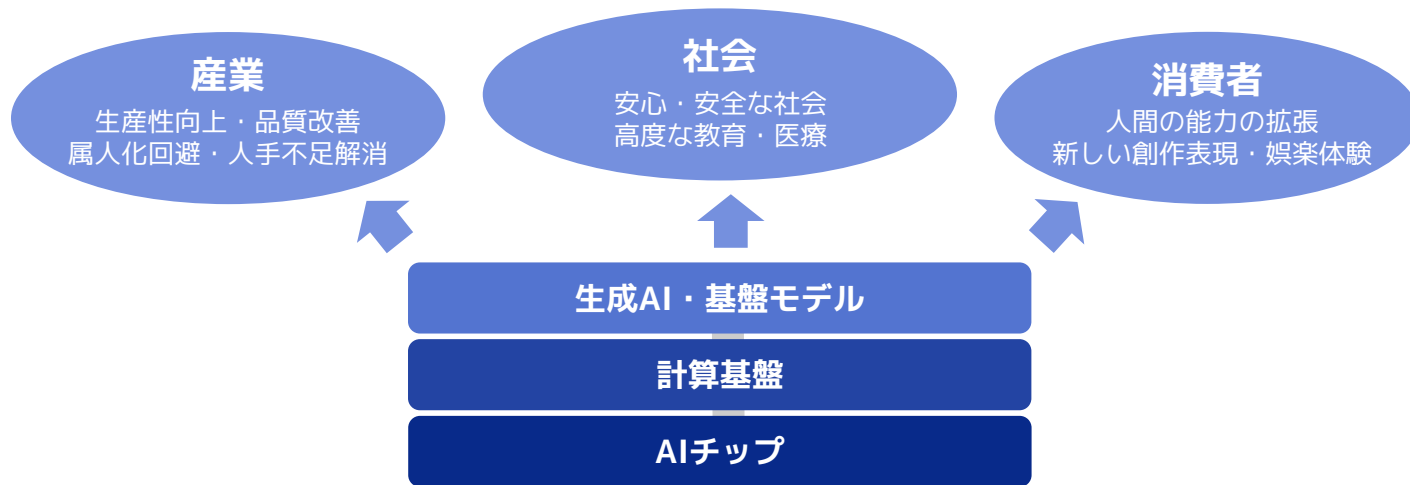
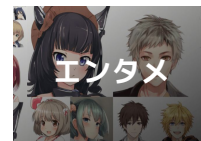
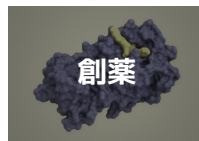
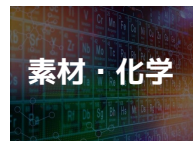
生成AI (推論) 向け
MN-Core L1000
(2026年提供予定)



MN-Core
次世代

PFNの事業: AI技術の水平展開

PFNは、AI技術のバリューチェーンを垂直統合し、産業、コンシューマー、社会に向けて様々な領域でソリューション・製品を水平展開しています。





ソリューション
・製品

PreferredAI™

生成AIを活用したプロダクト・サービス群

生成AIを活用し、想定ユースケースごとにパッケージ化したプロダクト・サービス群をPreferredAI（プリファードAI）として提供しています。また、生成AIに関するあらゆる技術要素を利用し、お客様の課題・ニーズに合わせた独自ソリューションを提供します。

PreferredAI

PreferredAI Slide Review



スライドを自動レビューし問題点をハイライト。サマリーも出力します。

PreferredAI Talent Scouter



AIアバターで対話実務を再現。接客練習や評価を支援します。

PreferredAI Notes



蓄積された自社のドキュメントや検索データを元に数秒でノートを生成。

PreferredAI Automation



あらゆる定型業務をLLMでミニアプリ化して再利用・効率化。

PreferredAI Insight Scan



口コミやアンケート、日報など、大量の文章を賢く分類してサマライズします。

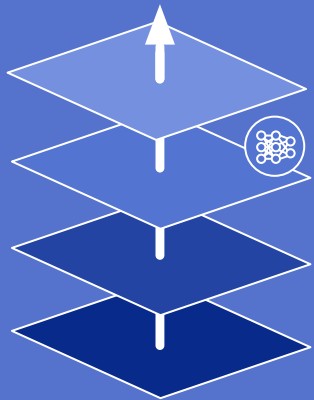
目次

Agenda

- 1 会社概要
- 2 PFNのAIガバナンス体制
- 3 過去の機械学習サービスの品質管理事例
- 4 まとめ



生成AI・ 基盤モデル



PLaMo™

PFNグループの生成AI基盤モデル

PLaMoは、PFNのグループ会社Preferred Elements (PFE)がフルスクラッチで開発する純国産の生成AI基盤モデルです。高品質な学習データを用い、既存の大規模言語モデルをベースに用いずに全く新規に開発。主要な日本語ベンチマークで世界最高レベルの精度を示しています。



- 商用版**PLaMo Prime**を2024年12月にリリース
- エッジ向け小規模言語モデル **PLaMo Lite** を提供中
- 金融・医療などの特定タスクに強いモデルも順次リリース予定

世界最高クラスの日本語性能

- 主要な日本語ベンチマーク*でGPT-4を超える精度を記録
- 現在も精度向上中
- 日英の翻訳性能にも優れる

*日本語LLMの性能を測るため標準的に使われるベンチマーク「Jaster」を使用

純国産フルスクラッチモデル

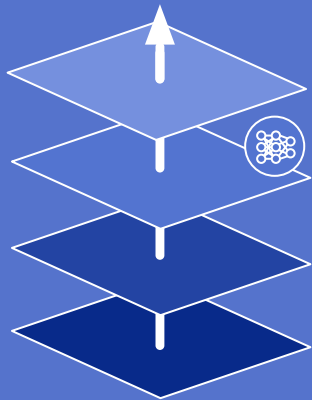
- 独自のアーキテクチャと学習データでゼロから事前学習・指示学習を実施
- 他社モデルベースではないため社外ライセンスの縛りや開発上で不明瞭な点がない

APIで簡単に導入可能

- API経由でクラウドを介して提供、簡単なコードの書き換えで実装可能
- モデル自体をお客様のクラウドなオンプレミス環境に実装することも可能



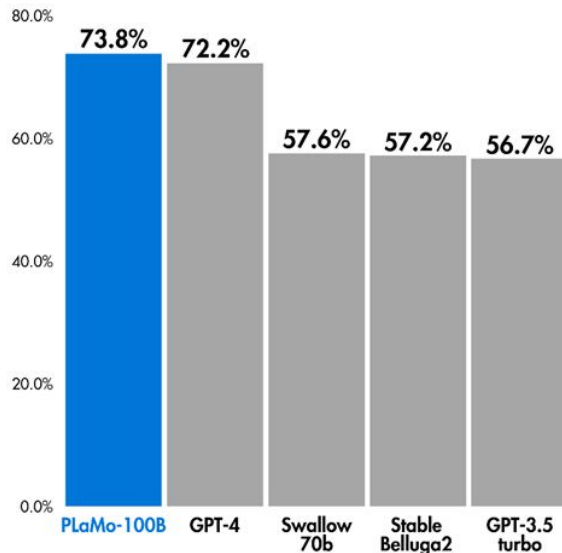
生成AI・ 基盤モデル



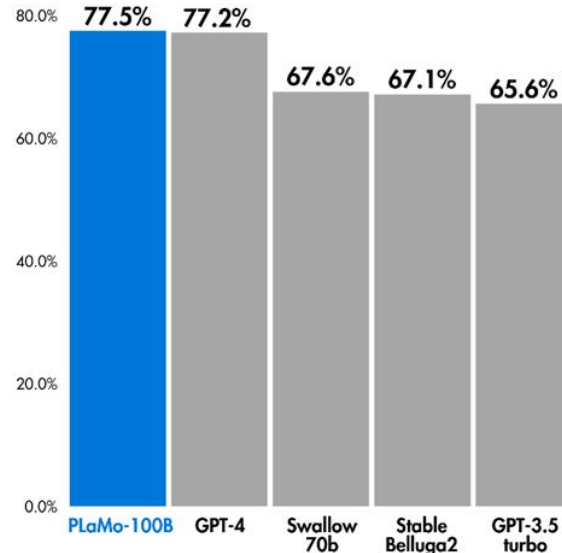
PLaMo™ β版

日本語性能世界トップレベルの大規模言語モデル

Jaster 0shot



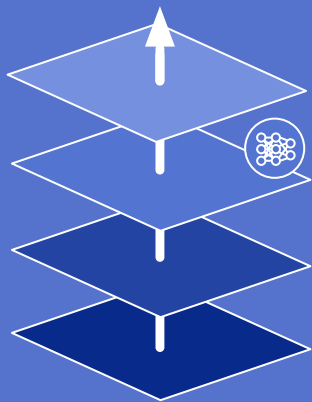
Jaster 4shot



※Jaster：日本語LLMの性能を測る標準ベンチマーク
※これらの数値はGENIAC内での評価値



生成AI・ 基盤モデル

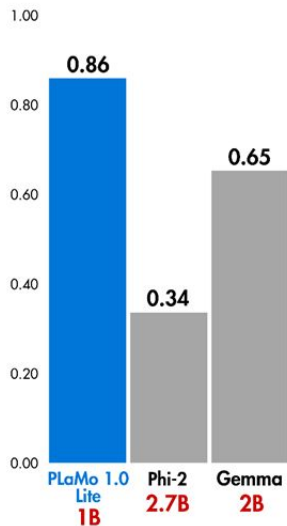


PLaMo™ Lite

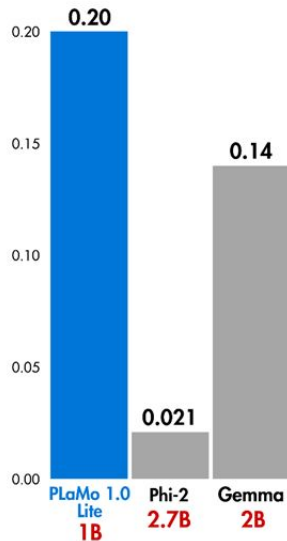
エッジデバイス向けの小規模言語モデル

- エッジデバイスで動かせる10億パラメータの小規模言語モデルの提供開始
- PLaMo-100Bの開発で培った技術に応用
- 同程度のサイズのモデルと比べて世界最高レベルの日本語性能を達成

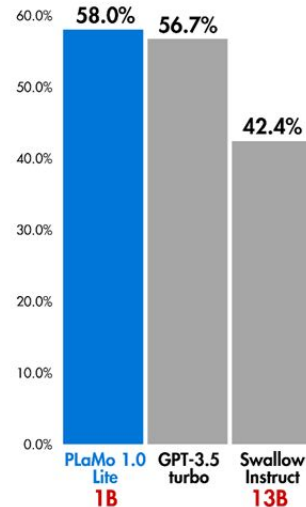
JSQuAD F1
(日本語質問応答)



WMT20 4shot
(日英翻訳 En-Ja)



Jaster 0shot



PLaMo-13B

- plamo-13b (2023.09.25) : Apache License 2.0
- plamo-13b-instruct (2023.10.26) : Apache License 2.0
- plamo-13b-instruct-nc (2023.10.26) : CC-BY-NC-4.0

PLaMo-100B

- plamo-100b (2024.9.24) : PLaMo非商業用ライセンス契約

PLaMo Prime (商用版)

- (2024.06.14) : PLaMo API、PLaMo Chat利用規約

PLaMo2

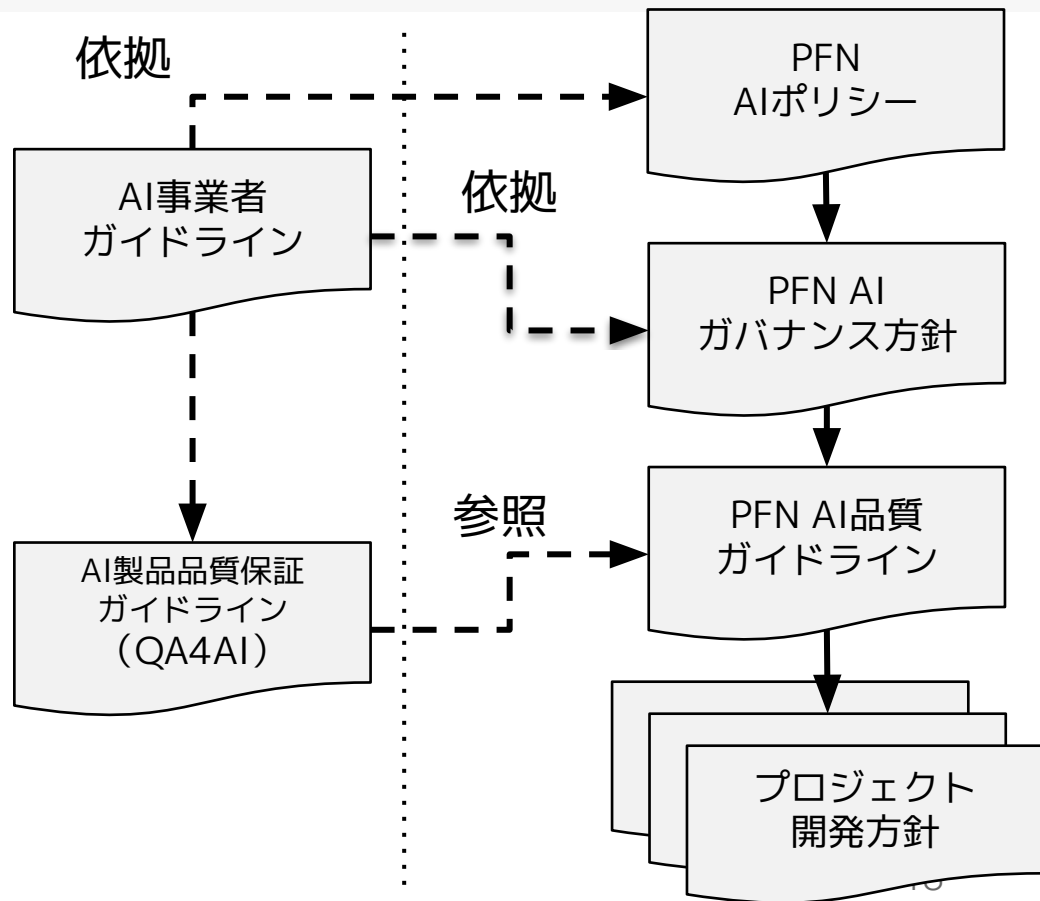
- plamo-2-1b (2025.2.6) : Apache License 2.0
- plamo-2-8b (2025.2.21) : PLaMoコミュニティライセンス契約

リリース時にリスク評価(ERM)を実施

中分類	小分類	項番	リスク項目
1. 技術領域	a 技術領域(規制、コンプライアンス、その他)	1	医療関係・BIO関係
		2	金融関係
		3	化学
		4	交通
		5	規制のある領域(輸出入規制を除く)
		6	その他
	b 人命等に関わる事故の可能性	7	アクチュエーター なし/小規模/大規模
		8	事故の可能性 なし/小規模/大規模
	c 環境汚染	9	通常時に懸念がある
		10	事故時に懸念がある
	d 倫理面と社会的な影響(人権・公平性)	11	社会的バイアスのかかった出力を行う可能性がある(犯罪予測、内定判断、保険、Tay、動物判定 google など)
		12	
	a データ・情報・	1	個人情報扱う(氏名、住所等、画像、医療情報、行動履歴)
		2	広義の個人情報扱う(Cookie など: GDPR)
		3	機微情報を扱う
		4	インターネットサービスを提供する

→ 大規模言語モデルの開発に伴う、AIガバナンス体制強化の必要性を認識

規程・ガイドライン体系



PFNのAIに関する基本方針

AIに関する適切な企業活動
のための体制・規程

プロジェクトが守るべき
ルールや手続き

ガイドラインの対応可否・
選択理由などを記述

1. 透明性（正しく伝える）

私たちは、その技術自体が何をするか、何ができるか、何ができないかを可能な限り明確にします。それがどのような仕組みで動くかを、論文その他の出版、ブログ、ドキュメント等を通して、様々なレベルのステークホルダーが理解できるよう、丁寧に説明します。

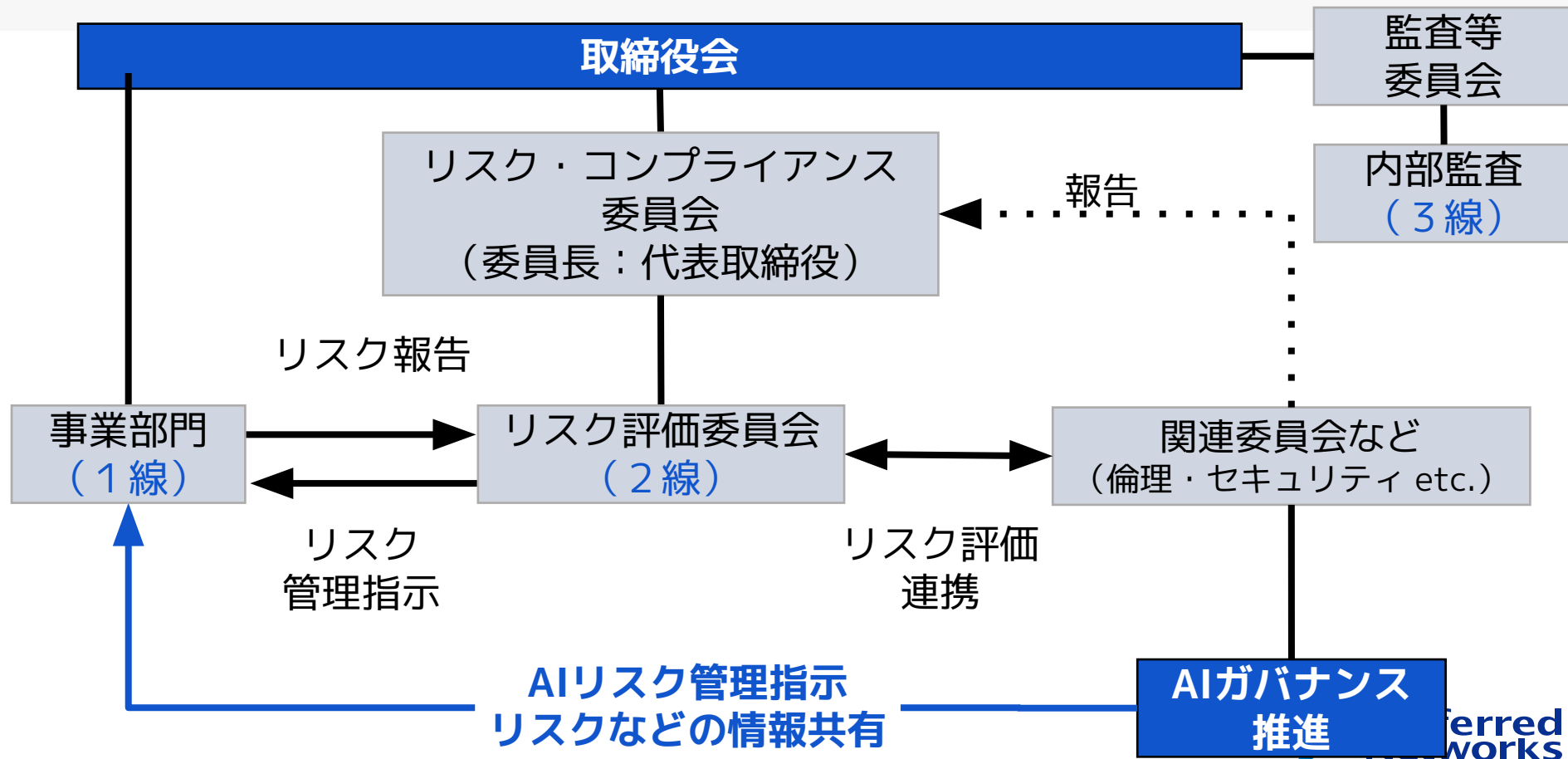
2. リスク（適切に怖がる）

私たちは、その技術を使うことによって、社会の様々な場面で起きうるリスクを、私たちの想像力が及ぶ限り考慮します。リスクには、定量化できないもの、想定しにくいものがあり、また、ある時点での社会通念ではリスクではなかったものが、その後、にリスクと考えられるようになることもあります。新技術は常に便益とリスクを伴います。私たちは、社会との対話を通して、そのようなリスクに対応していく努力を惜しみません。

3. インテグリティ（見たくないものを、見る）

新技術の開発や利用にあたっては、時として私たちに都合の悪い発見があるかもしれません。私たちは、そのようなものから目を背けることなく、誠実に対応します。

規程・ガイドライン体系



開発方針(抜粋)

【モデルケース】プロダクトA開発方針

XXXX年X月XX日

文責:XXXXX(<メールアドレス>)

更新情報

- 202X年X月XX日:初版作成
- 202X年X月XX日:更新作成
- 202X年X月XX日:更新作成

本ドキュメントについて

PFNのプロジェクトの管理者は、プロジェクトでのAI品質管理方法を定めた開発方針を策定しなければならない。本ドキュメントは、各プロジェクトが開発方針を定めるために利用可能な開発方針テンプレートである。「PFN AI品質ガイドライン」には各プロジェクトがAIシステムの開発・提供を行うにあたって、遵守すべき一般的な内容についてまとめている。必ずしも、上記ガイドラインのすべての項目を満たす必要はないが、遵守事項を採用しない場合は、開発方針にその理由とリスクの担保方法を明確にすることが求められる。

開発方針(抜粋)

1.2. プロジェクト体制・データ管理ポリシー

記載内容

- プロジェクト体制(責任者、開発メンバー、運用体制など)
- データ管理ポリシー

プロジェクト体制が現段階で確定していない場合は、その旨を記載してください。また、データ管理ポリシーを作成していない場合は、その旨記載の上別途セキュリティチームにご相談ください

補足:このプロジェクトに関することを誰に聞けばよいか、どのデータに誰がアクセスすることができるのかを把握できるようにしてください

プロジェクト体制

- 責任者:担当VP xxxxx@
- メンバー: xxxxx@ xxxxx@ xxxxx@ xxxxx@
- 運用体制:データ管理ポリシーに記載

データ管理ポリシー:(資料削除)

開発方針(抜粋)

2. 品質要素のチェックリスト

以下の表の左側のチェック項目に対して右側に回答を記入してください。本チェックリストは様々な種類・様々な開発段階の製品を網羅するように作成しているため、チェック項目すべてを満たすことを要求するものではありません。そのため、対応していない箇所はその旨を記載してください。

各項目の詳細は、AI品質ガイドライン「2. AI製品の品質要素」を参照してください。

補足: QA4AIは、監査的な視点よりもAI関連のシステム開発者の視点で書かれており、これまでのAI関連のシステム開発にかかわる、失敗談や成功事例に基づいた知見に基づく開発へのサジェッション・推奨事項が含まれていません。

2.1. Data Integrity: データの適切な管理など

(a) 学習データの量の十分性

(a.i) 想定する学習手法の適用前提や統計的観点から十分な量のデータがあるか。

想定回答例:

- 適用前提から十分と言えるデータ量を用意した
- 統計的モデルから十分と言えるデータ量を用意した
- 他の研究や過去の実績から十分と言えるデータ量を用意した
- POC的なプロジェクトであるため、評価を行いながらプロジェクトを進める

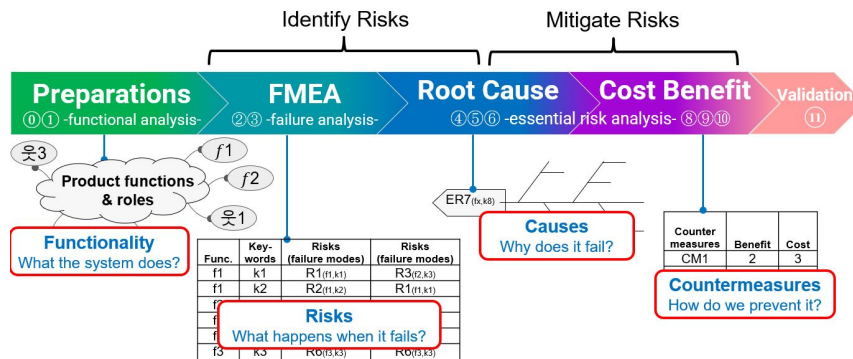
目次

Agenda

- 1 会社概要
- 2 PFNのAIガバナンス体制
- 3 過去の機械学習サービスの品質管理事例
- 4 まとめ

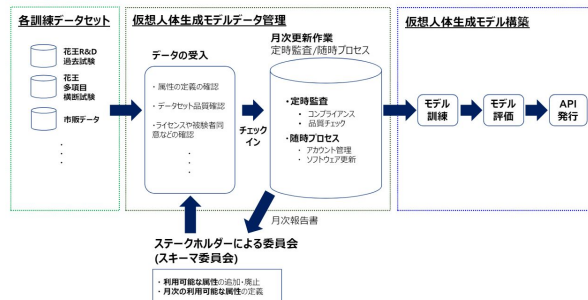
「実務で使われている機械学習」の品質管理の例

1. 産業機械における機械学習の品質管理



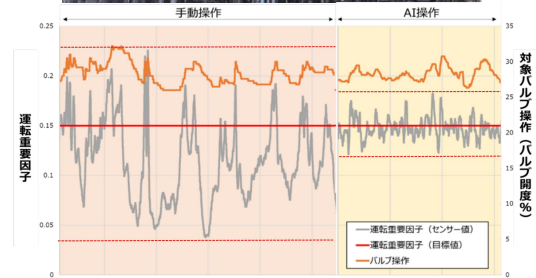
出典：G. Dominguez, "Quality Assurance for ML Devices A Risk Based Approach, APSEC 2023.

2. 仮想人体生成モデルの品質管理



出典：尾藤宏達ほか: 仮想人体生成モデルにおける品質管理, 日本ソフトウェア科学会第40回大会講演論文集, Tokyo, 2023.

3. 石油化学プラント自動運転の品質管理



出典

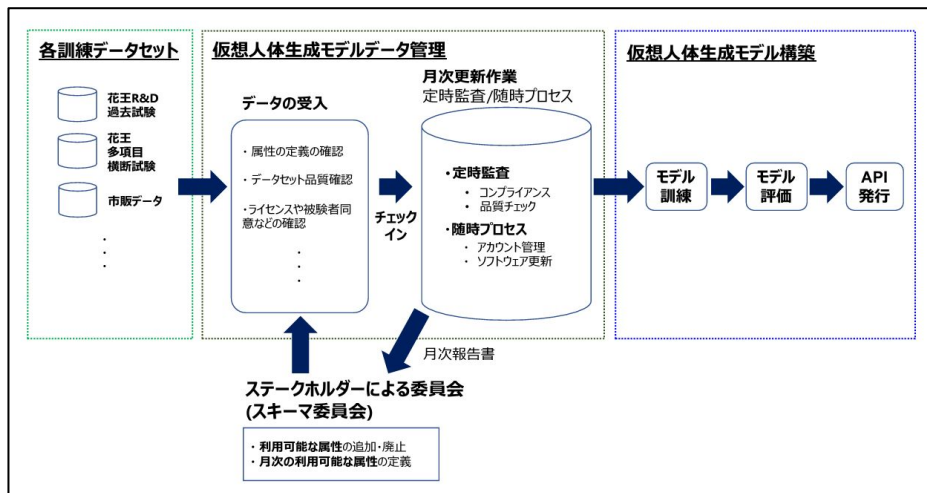
： <https://www.preferred.jp/ja/news/pr20230731/>、
 平井太一朗ほか: AIによるプラント自動運転の実証と信頼性評価について, 第54回安全工学研究発表回講演予稿集, Tokyo, 2021.

※図は丸山宏氏のスライドから作成

開発プロセスの品質管理の例：仮想人体生成モデル

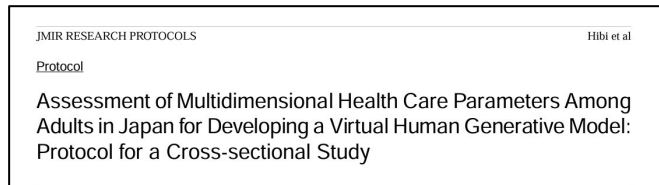
Product/Process/Performanceを様々な観点から評価

- ・ 毎月のデータセット払い出しの見直し
- ・ データセット・統計モデルの技術文書を公開
- ・ テストデータセットを開発プロセスから独立して保持
- ・ プライバシー漏えいリスク評価



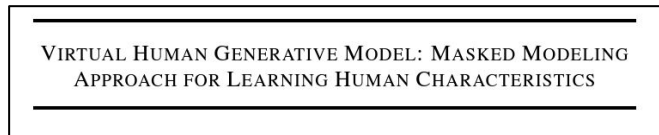
左図：尾藤宏達ほか: 仮想人体生成モデルにおける品質管理, 日本ソフトウェア科学会第40回大会講演論文集, 2023.

データセットプロトコル論文



Hibi et al., JMIR Res Protoc 2023;12:e47024

統計モデルホワイトペーパー



Oono et al., arXiv:2306.10656

従来の機械学習に対する品質管理の事例は増えつつある。一方で、大規模言語モデルを想定した品質管理は模索が続いている

目次

Agenda

- 1 会社概要
- 2 PFNのAIガバナンス体制
- 3 過去の機械学習サービスの品質管理事例
- 4 まとめ

本講演のまとめと今後の取り組み

本講演のまとめ

- 大規模言語モデルの開発における、固有のAIガバナンス体制強化の必要性
- 弊社のAIガバナンス体制構築の事例紹介
- 従来の機械学習に対する品質管理の事例は増えてつつあるが、大規模言語モデルを想定したAI品質管理は模索中

今後の取り組み

- AIレッドチーミング体制の拡充
- 社員向けチュートリアル
- プロダクト開発方針の事例拡充
- 外部機関・事業者との連携



Making the real world computable