プライバシー保護データ解析技術の社会実装

CREST：イノベーション創発に資する人工知能基盤技術の創出と統合化

2020.9.24

研究代表者：花岡 悟一郎（産総研）
主たる共同研究者：盛合 志帆（NICT） ※共同提案者
主たる共同研究者：浅井 潔（東京大）
主たる共同研究者：小澤 誠一（神戸大）
主たる共同研究者：菅原 貴弘（エルテス）

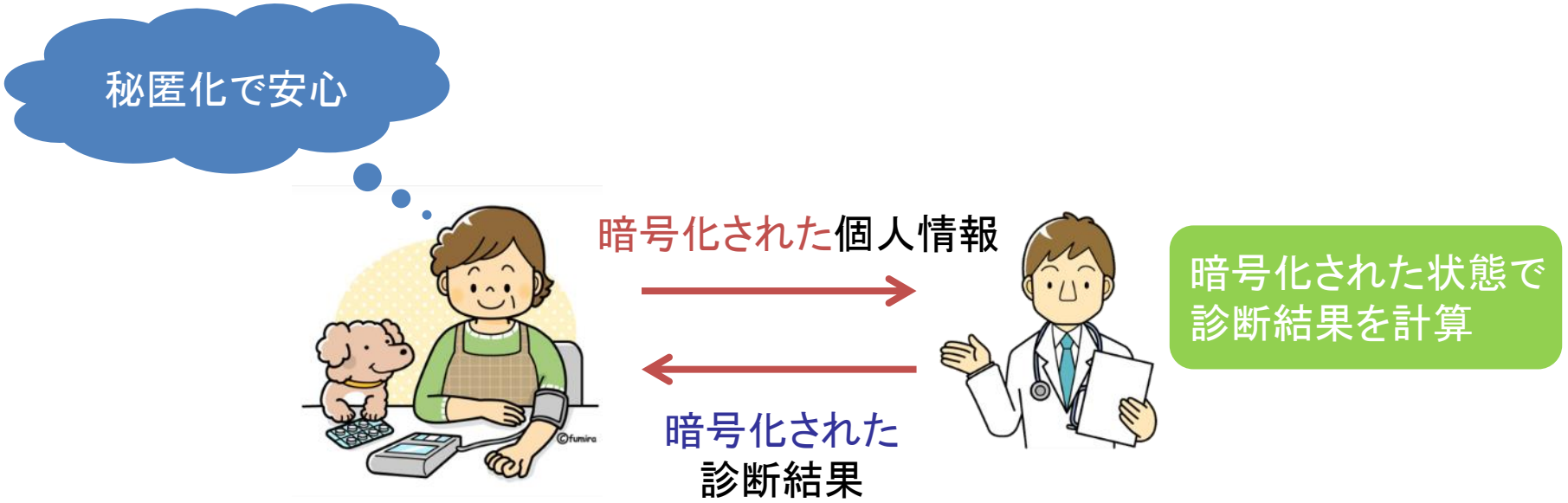
【背景】プライバシー侵害の懸念

- 人工知能を機微情報に適用することで、個人ごとにサービスが提供される「優しい社会」が実現される。
- しかし、プライバシー侵害の懸念がある。

オンライン自動健康診断サービスの例

【背景】秘匿計算技術

- 情報を秘匿したまま解析を行う秘匿計算技術の研究開発が進んでいる。



AI-CREST スモールフェーズ

花岡　悟一郎

安全な秘匿化データ処理を実現する汎用依頼計算技術

研究者

花岡　悟一郎

産業技術総合研究所
情報技術研究部門
研究グループ長

主たる共同研究者

浅井　潔	東京大学　大学院新領域創成科学研究科　教授

研究概要

情報漏洩の心配のないサービスを、誰でも、いつでも、圧倒的低コストで提供可能とすることを目指します。本研究では、最先端暗号技術の統合によりサービスごとの個別設計を不要とする汎用的秘匿化技術を創出し、依頼に応じて秘匿化データ処理を代行する汎用秘匿化依頼計算システムの開発を行います。人工知能に基づく自動健康診断をはじめとする、個人ごとにきめ細かなサービスが提供される優しい社会の実現に貢献します。

盛合　志帆

複数組織データ利活用を促進するプライバシー保護データマイニング

研究者

盛合　志帆

情報通信研究機構
サイバーセキュリティ研究所
室長

主たる共同研究者

小澤　誠一	神戸大学　数理・データサイエンスセンター　教授
菅原　貴弘	株式会社エルテス　代表取締役

研究概要

複数の異なる業種・組織が有する実社会の膨大なデータを統合して利活用する際に、プライバシー保護やデータ機密性の確保が課題となっています。本研究課題では、暗号技術や人工知能技術を活用し、プライバシーを保護した状態で高速にデータ分析や異常検知を行う技術の研究開発を行います。この技術を金融分野における不正送金検知や顧客に合わせた金利決定の支援に応用し、フィンテックにおけるイノベーション創出を目指します。

スモールフェーズにおける両チームの課題認識と方針

最終目標は
もともと共通！

花岡課題

プライバシー保護データ解析技術の
ビジネス展開を阻害しているのは、
汎用性の欠如による膨大な開発コスト
⇒汎用的データ処理秘匿化技術の開発

社会展開元となる
企業と連携し、
商用システムとし
て作りこむ

盛合課題

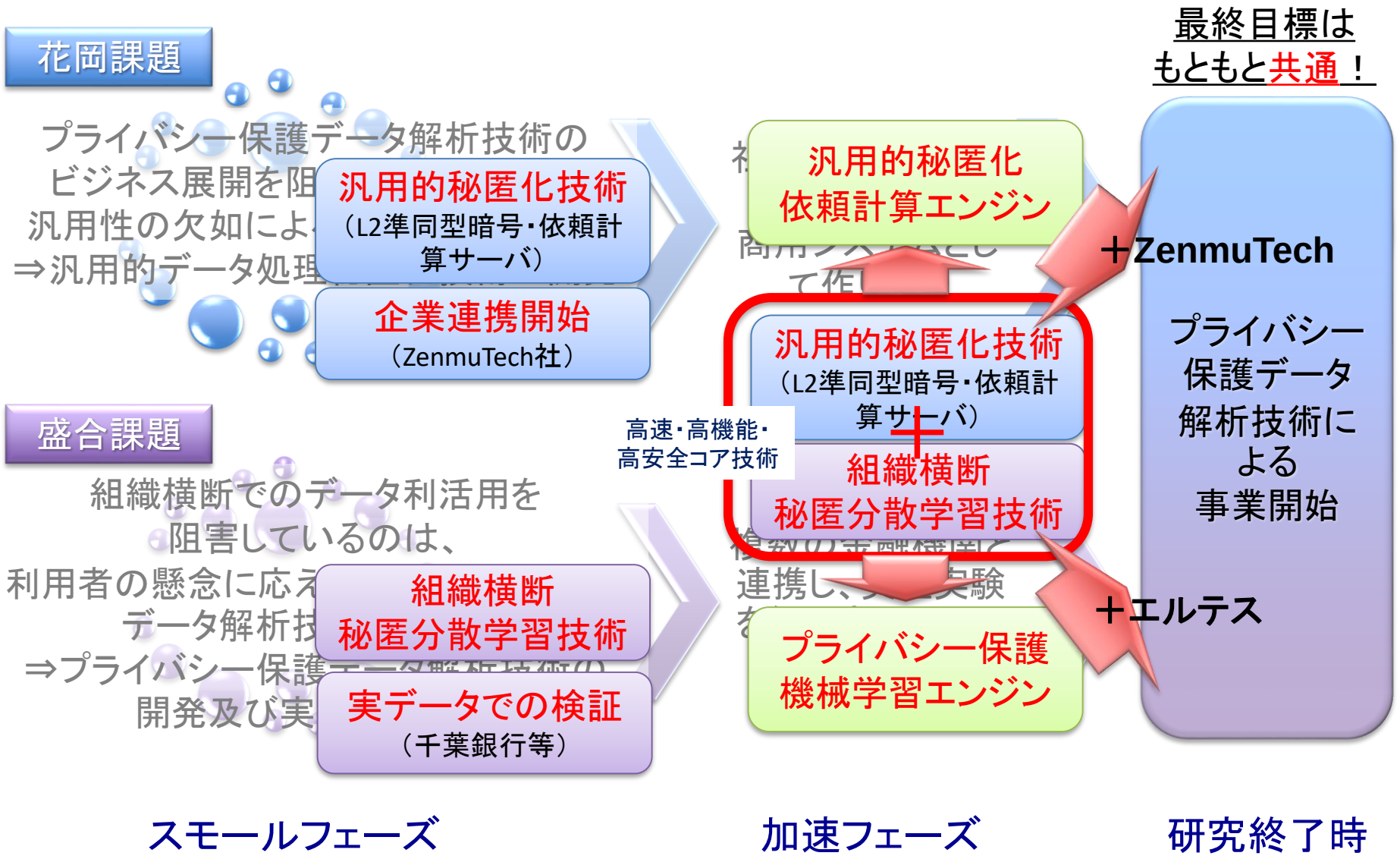
組織横断でのデータ利活用を
阻害しているのは、
利用者の懸念に応えるプライバシー保護
データ解析技術の未浸透
⇒プライバシー保護データ解析技術の
開発及び実社会での実証

複数の金融機関と
連携し、実証実験
を経て実システム
を開発

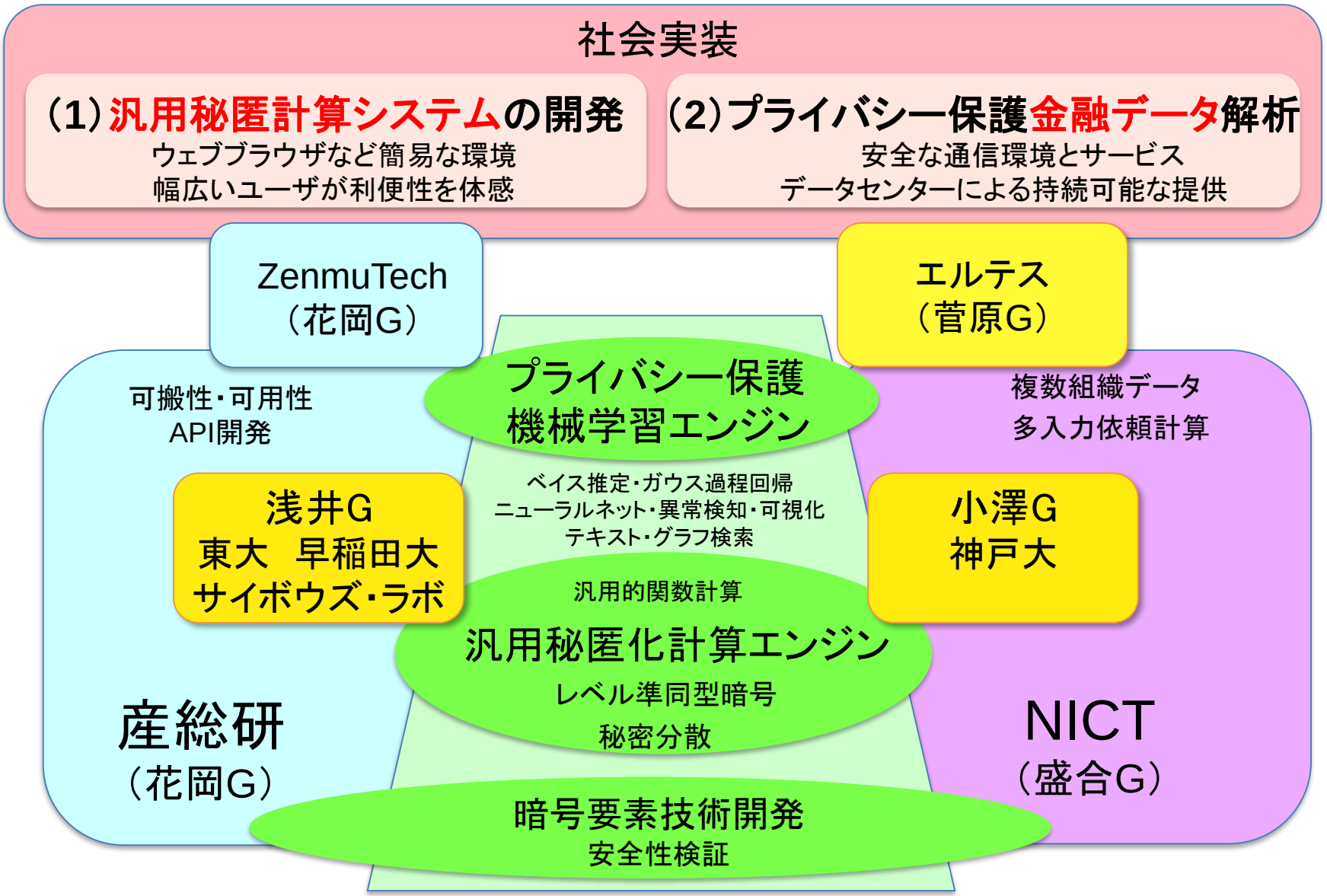
プライバシー
保護データ
解析技術に
よる
事業開始

スモールフェーズ　　加速フェーズ　　研究終了時

加速フェーズにおける研究開発方針

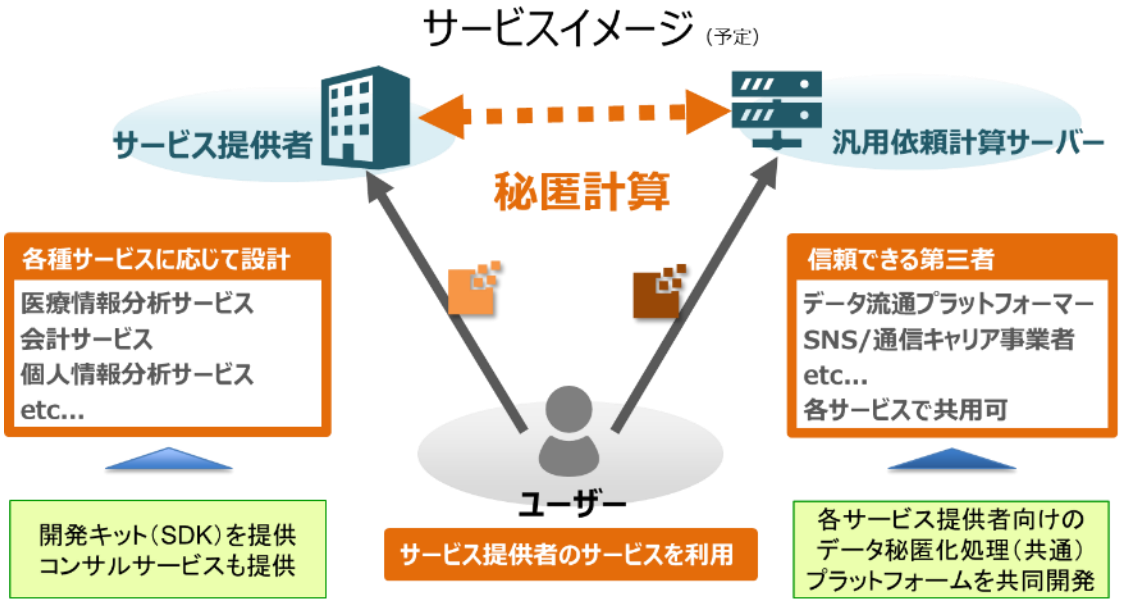


研究体制


社会実装
(1) 汎用秘匿計算システムの開発
ウェブブラウザなど簡易な環境
幅広いユーザが利便性を体感
(2) プライバシー保護金融データ解析
安全な通信環境とサービス
データセンターによる持続可能な提供
ZenmuTech
（花岡G）
エルテス
（菅原G）
プライバシー保護
機械学習エンジン
可搬性・可用性
API開発
複数組織データ
多入力依頼計算
ベイス推定・ガウス過程回帰
ニューラルネット・異常検知・可視化
テキスト・グラフ検索
浅井G
東大　早稲田大
サイボウズ・ラボ
小澤G
神戸大
汎用的関数計算
汎用秘匿化計算エンジン
レベル準同型暗号
秘密分散
産総研
（花岡G）
NICT
（盛合G）
暗号要素技術開発
安全性検証

研究開発内容(1)

商用利用可能な汎用秘匿計算システムの開発



- アルゴリズムの理論的高速化
  - ✓最新の知見をもとに洗練化
  - ✓数学的安全性証明も
- 高速実装(アセンブラ実装等)
- 通信インターフェース
- 実サービスを提供可能な依頼計算サーバの設置
- ライブラリ・API化
  - ✓汎用秘匿化依頼計算エンジン
  - ✓専門的研究者でなくても利用可能に
- サンプルアプリケーションの実装
  - ✓事業展開の際の顧客向けサンプル

研究開発内容（1）

社会実装を実現するため三者で共同検討中

㈱野村総合研究所

社会課題への適用

- 社会課題への適用ニーズ探索
- 秘匿計算の適用性検証
- 適用ソリューション開発

産総研

理論

- アルゴリズムの高速化・高機能化
- 安全性検証
- 論文執筆・学会発表

㈱ZenmuTech

開発

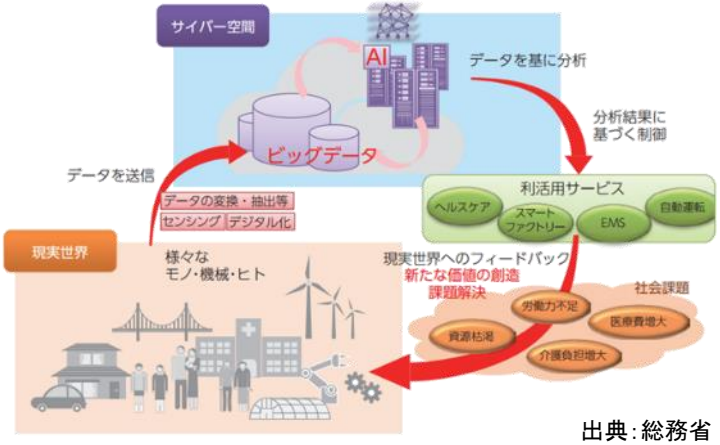
- エンジン（ライブラリ）開発
- 秘匿計算機能実装
- 論文執筆・学会発表（共同）

研究開発内容（1）

社会実装を実現するため三者で共同検討中

NRI　AIST　ZENMU TECH

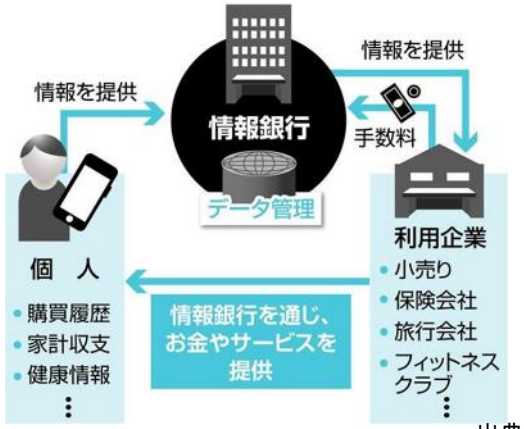
ビッグデータ利活用に対するセキュリティ（プライバシー保護）価値の付与



出典：総務省

- アナリティクス基盤サービス
- 秘匿計算環境パッケージ（クラウド）
- 金融業界向けサービスへの適用

パーソナルデータ利活用に対するセキュリティ基盤としての活用



出典：産経新聞

- 情報銀行を介したプライバシー情報の授受・提供の過程において秘密分散ならびに秘匿計算技術を活用したデータ管理基盤を構築
  これにより情報銀行はデータの「やりとり」はするがその中身は一切「知ることなく」高セキュアな信託機能を発揮することができる
  ⇒セキュアPDSサービス

研究開発内容(1)

ソリューション

外部リソースを安全に活用できるデータ処理基盤

外部にデータ復元ポイント無し ← 秘匿計算技術

外部からのデータ操作が可能になり
人材の配置が自由に

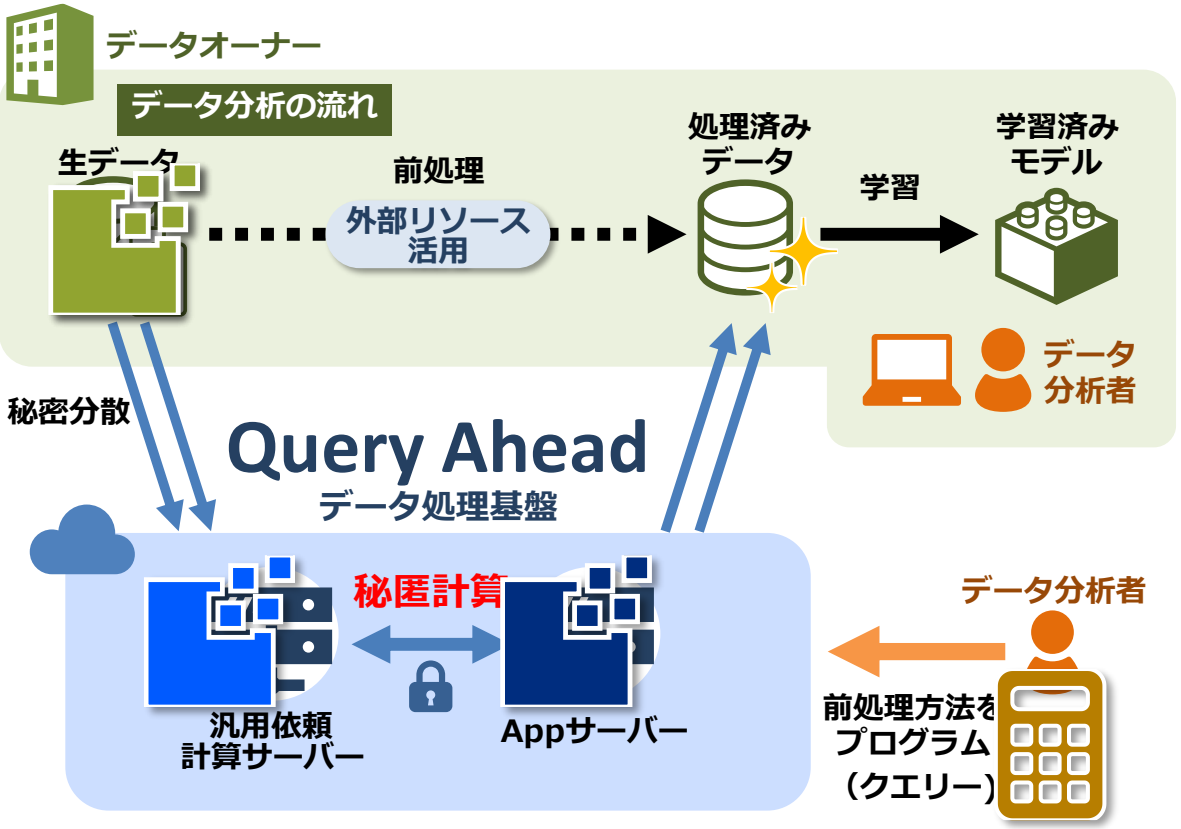
外部計算リソースの活用により
大規模データの利活用が容易に

リソースの解放によるコスト削減＋効率化

研究開発内容(1)

処理の流れ

1. データは秘密分散によって秘匿化
2. 秘密分散されたデータが汎用依頼計算 (CoSC) サーバーとAppサーバーへ送信される
3. データ分析者によってプログラムされた前処理を実行 ← 秘匿計算
4. それぞれのサーバーから処理結果をデータオーナーが収集し復元



研究開発内容(1)

クエリー

■ **データ分析者が慣れ親しんでいるSQL的な操作が可能**

- 基本演算 (加算、減算、乗算)
- SELECT句 (列の選択、列同士の演算)
- WHERE句 (等価評価による抽出)
- FROM句 (テーブルの選択)

今後さらに実行可能処理を増強予定

■ **数行の記述でクエリー発行可能**

```
q = (Query('result')
     .select(date,
             kari_code,
             kashi_code,
             kari_amount,
             kashi_amount,
             (kari_code + kashi_code) * 10)
     .where(date == _date('2019/8/27'))
     .from_('journal'))

app_client.req_query(q)
```

SELECT句

WHERE句

FROM句

基本演算の記述

日付に一致する行を抽出

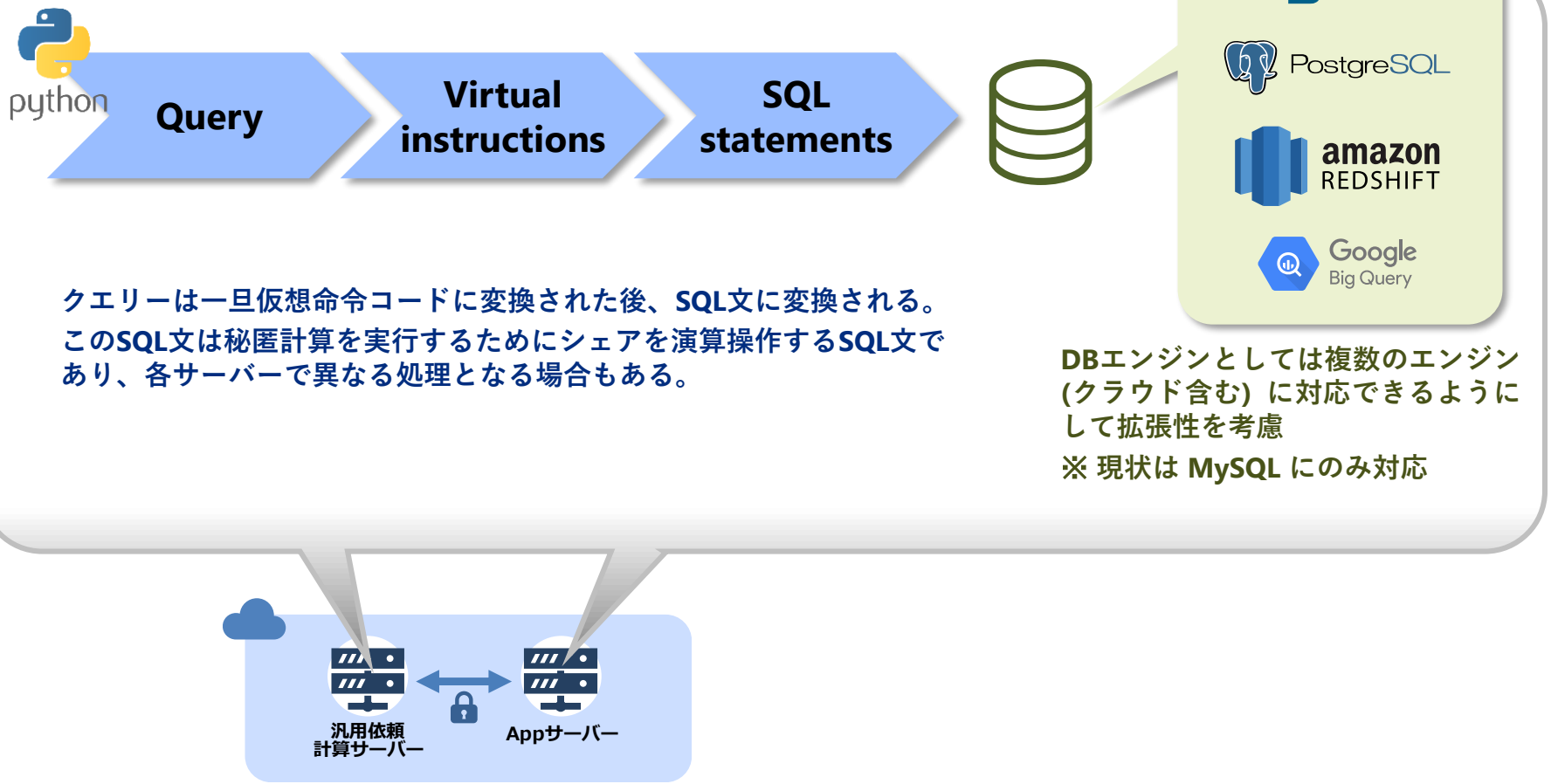
研究開発内容(1)

大規模データの扱い

■ シェアの演算処理は既存のDBを活用

- メモリに乗り切らないような大規模データの演算はDBに任せる（餅は餅屋）
- 限られた開発リソースで、高速に開発を推進するため

■ 秘匿計算処理をSQL文に翻訳するエンジンを開発



研究開発内容(1)

内部処理の例

Query

```
q = (Query('result')
     .select(kari_code + kashi_code)
     .from_('journal'))
```

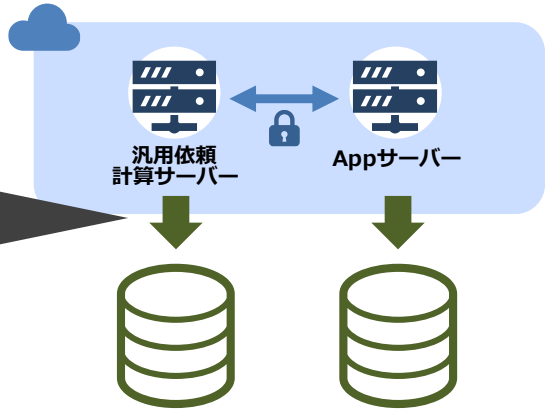
Virtual instructions

```
add r0, kari_code, kashi_code
select tblresult, r0
```

秘匿計算を実行するためのSQL文

SQL statements

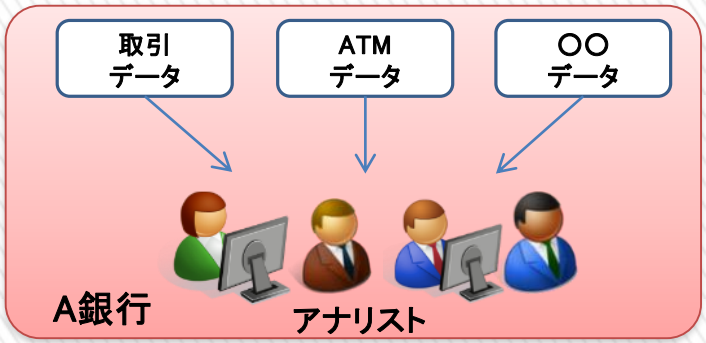
```
CREATE TEMPORARY TABLE r0 SELECT
  tbljournal._id,
  CAST((CAST((kari_code) AS DECIMAL(11))+
     CAST((kashi_code) AS DECIMAL(11))+
     4294967296
    ) % 4294967296 AS UNSIGNED) AS r0
  FROM tbljournal;
CREATE TABLE tblresult SELECT
  tbljournal._id, (r0) FROM tbljournal
  INNER JOIN r0 ON tbljournal._id = r0._id;
```

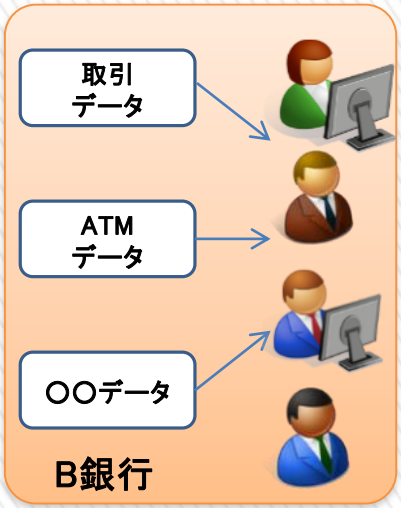


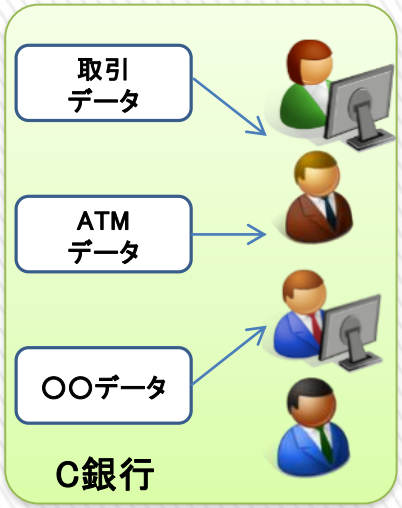
各サーバーが持つDBに対して
生成SQLを実行

目標:金融分野でのデータ統合利活用

現状

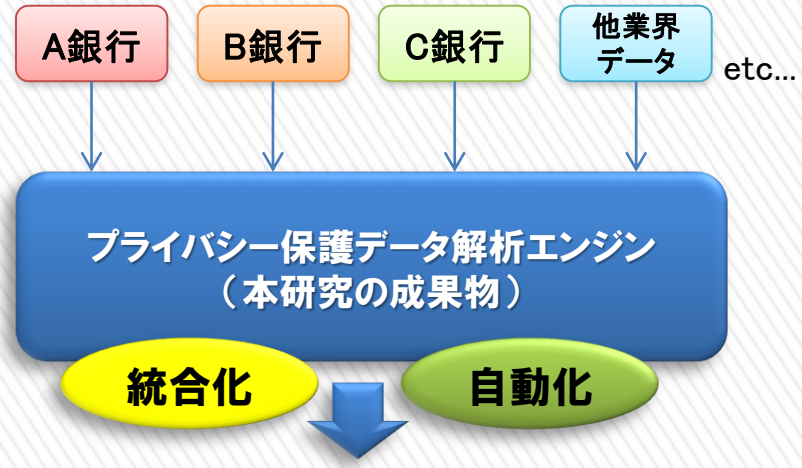






個々の金融機関内で分析
- コストや精度に課題
- AIを導入しようにもデータ量不足

めざす構想



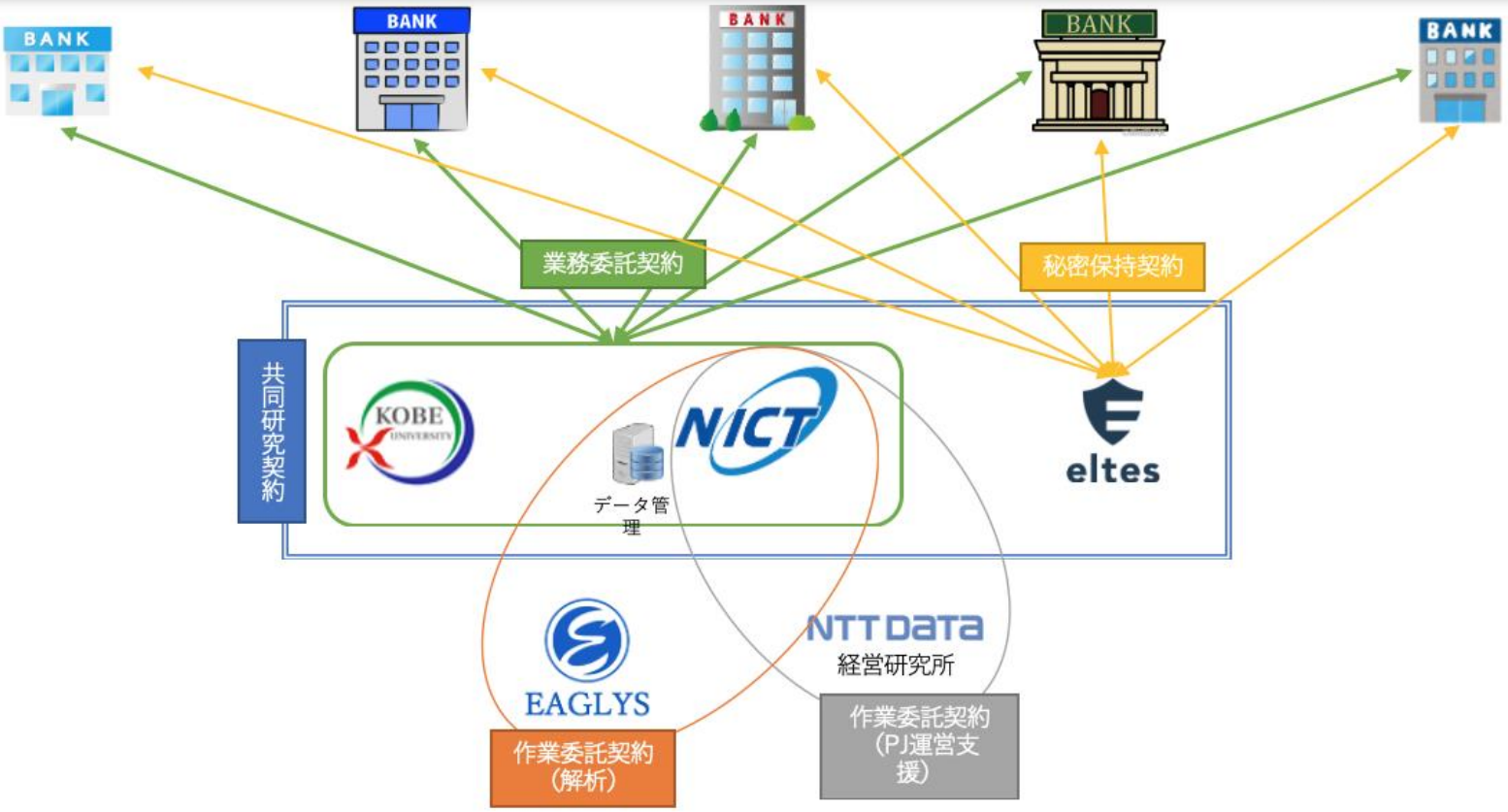
不正取引の検知,
与信管理, マーケティング

- 調査コスト削減
  - ➢疑わしい取引のスクリーニング、自動検知
- 調査属人化の回避
- 調査精度の向上
  - ➢今まで見つからなかった検知が可能に！

全体体制

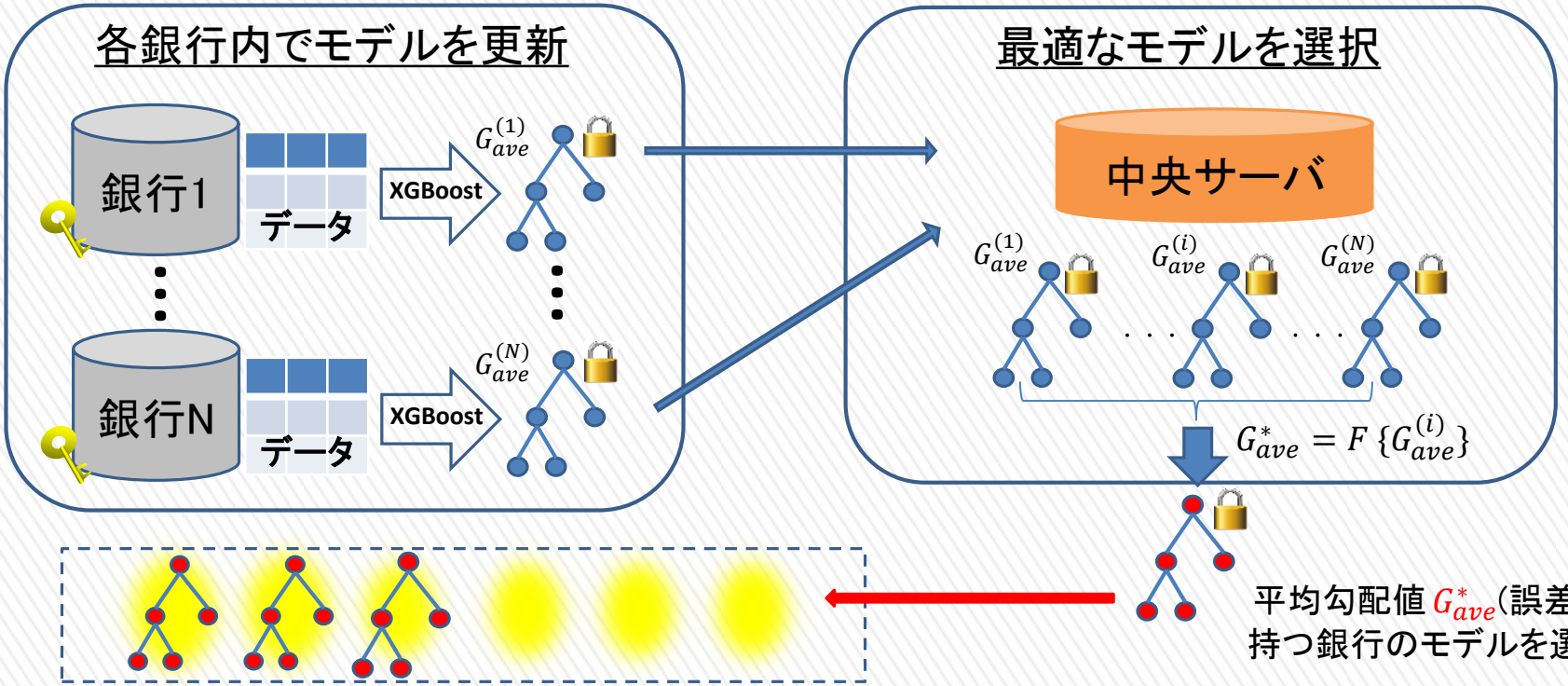
銀行5行と
実証実験中

千葉銀行, 三菱UFJ銀行, 中国銀行, 三井住友信託銀行, 伊予銀行が不正送金検知の実証実験に参加、オープンイノベーションによる実施体制を構築



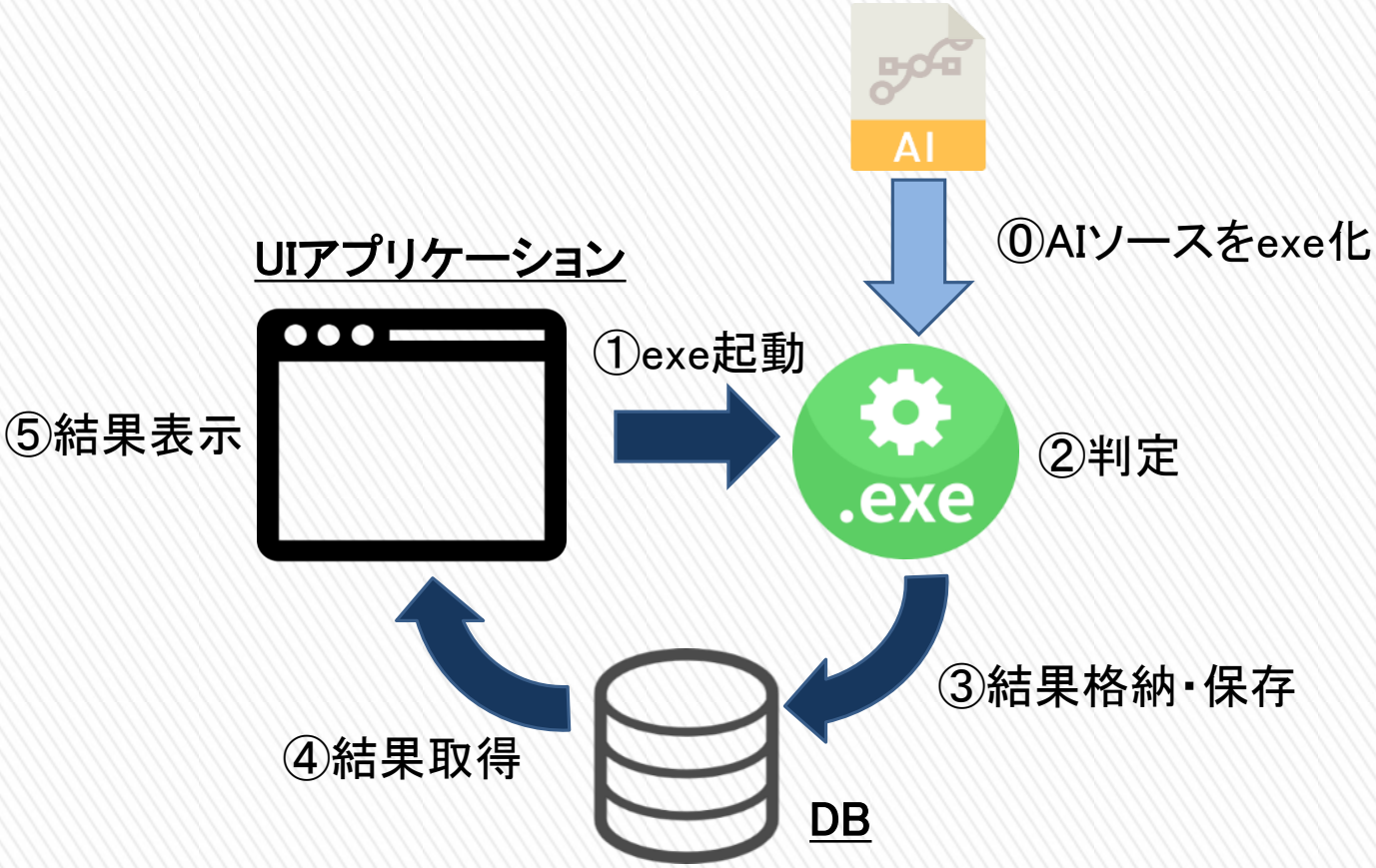
プライバシー保護型協調学習

- 組織が持つデータを外部に開示することなく機械学習を行うプライバシー保護学習システム
  - DeepProtect(深層学習をベースとした方式)
  - PP-FL-XGBoost(XGBoost をベースとした方式)
    - 出力に対する一定の説明性を持つモデルを学習



社会実装に向けた取り組み

- 現在、金融機関と連携の上、本格的な社会実装(事業化)に向けて金融機関ニーズ、業務要件、機能要件の調査中
- 銀行内稼働AI判定アプリケーションの実装を開始



課題解決がもたらすイノベーションの創出

- プライバシー保護データ解析技術に基づく
  ビジネス展開が可能に

- 人工知能に基づく、機微情報を用いた
  情報サービスの提供が加速

- 機微情報を用いたデータ社会の恩恵を、
  誰もが、簡単に、安全に享受

誰もが、簡単に、安全に → イノベーション