ZKP-DB 説明資料

プライバシーを守りながら医療データを最大限活用するシステム

☒ 目次

エグゼクティブサマリー

一言で言うと

「医療データを暗号化したまま分析・機械学習できるシステムです。個人情報は完全に保護され、データ漏洩しても中身は見えません。」

3つの特徴

1. **完全なプライバシー保護**: データは暗号化され、秘密鍵なしでは絶対に復号できません
2. **暗号化されたまま計算**: 統計分析や機械学習が、データを復号せずに実行できます
3. **データの正当性を証明**: ゼロ知識証明により、データが本物で改ざんされていないことを証明できます

対象顧客

- **データ提供者**: カルテ会社、病院、健康保険組合
- **データ購入者**: 製薬会社、研究機関、AI企業、保険会社

医療データ活用の現状と課題

現状: 貴重なデータが眠っている

> 病院・カルテ会社
> ├ 数十万〜数百万人の患者データ
> ├ 診断、治療、検査結果
> └ 新薬開発や医療研究に有用
>
> しかし...
>
> × プライバシー規制が厳しい
> × 患者の同意取得が困難
> × データ漏洩のリスク
> × 外部提供のハードルが高い
>
> → データが活用されず、社会的損失

社会的なニーズ

- **製薬会社**: 新薬開発のための臨床データが欲しい
- **研究機関**: 大規模な疫学研究をしたい
- **AI企業**: 診断支援AIを訓練したい
- **保険会社**: リスク評価の精度を上げたい

しかし、プライバシー保護と利活用の両立が困難

従来の仮名化の限界

仮名化とは

個人を特定できる情報（氏名、住所など）を削除・置換する方法

> 元データ:
> 氏名: 山田太郎
> 年齢: 45歳
> 性別: 男性
> 住所: 東京都渋谷区
> 疾患: 糖尿病
>
> ↓ 仮名化
>
> 仮名化データ:
> 氏名: [削除]
> 年齢: 45歳
> 性別: 男性
> 住所: 東京都
> 疾患: 糖尿病

☒ 仮名化の深刻な問題

問題1: 再識別のリスク

実際の研究結果:

- 「年齢 + 性別 + 郵便番号」だけで、**米国民の87%が特定可能**
- 複数のデータセットを組み合わせると、個人が特定される

> 例: 45歳、男性、東京都、糖尿病
>
> ↓ 他のデータと照合
>
> SNS: 45歳、男性、東京都在住
> 選挙人名簿: 45歳、男性、渋谷区
> 健康保険データ: 45歳、糖尿病治療歴
>
> → 個人が特定される！

問題2: データ漏洩時のリスク

仮名化データでも、**データ自体は平文**（暗号化されていない）

> サーバーがハッキングされると...
>
> 攻撃者:
> 「45歳、男性、東京都、糖尿病」
> → データの内容が丸見え
> → 他のデータと照合して個人特定

問題3: 法的なグレーゾーン

- GDPR（欧州）: 仮名化データは依然として「個人データ」扱い
- 規制対象のまま
- 国際的なデータ移転が困難

☒ ビジネス上の制約

仮名化では以下が困難:

- × 詳細な個別患者データの提供
- × 個別レベルの分析
- × 高度な機械学習モデルの訓練
- × グローバル展開

ZKP-DBの革新的なアプローチ

2つの先端技術の組み合わせ

1. 準同型暗号（Homomorphic Encryption）

普通の暗号化の問題:

> 暗号化 → 計算したい → 復号が必要 → セキュリティリスク

準同型暗号の革新:

> 暗号化 → 暗号化のまま計算 → 結果を復号 → 安全！

例えで説明:

> 透明な金庫に書類を入れる（普通の暗号化）
> → 読むには開けないといけない
>
> 不透明だけど操作可能な金庫（準同型暗号）
> → 開けずに中の書類を整理できる

2. ゼロ知識証明（Zero-Knowledge Proof, ZKP）

証明の革新: 「情報を明かさずに、その情報が正しいことを証明できる」

例: 洞窟の例

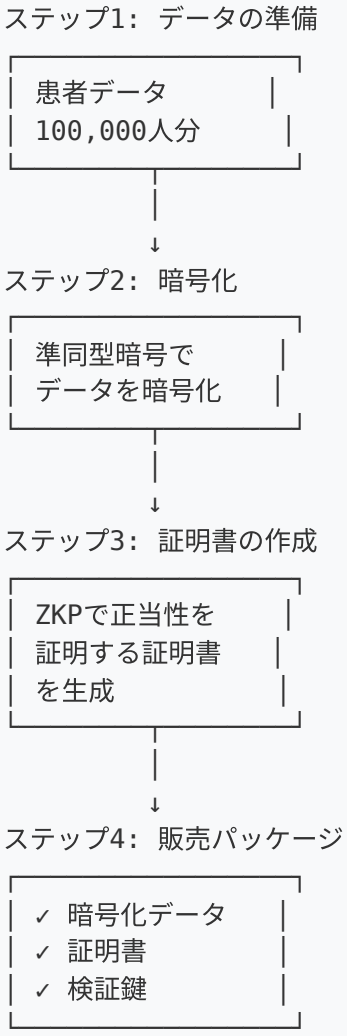
> あなた: 「この洞窟の秘密の扉を開ける鍵を持っている」
> 相手: 「証明して」
> あなた: [洞窟に入って反対側から出てくる]
> 相手: 「鍵を持っているのは確か。でも鍵そのものは見ていない」
>
> → 鍵を見せずに、持っていることを証明

医療データへの応用:

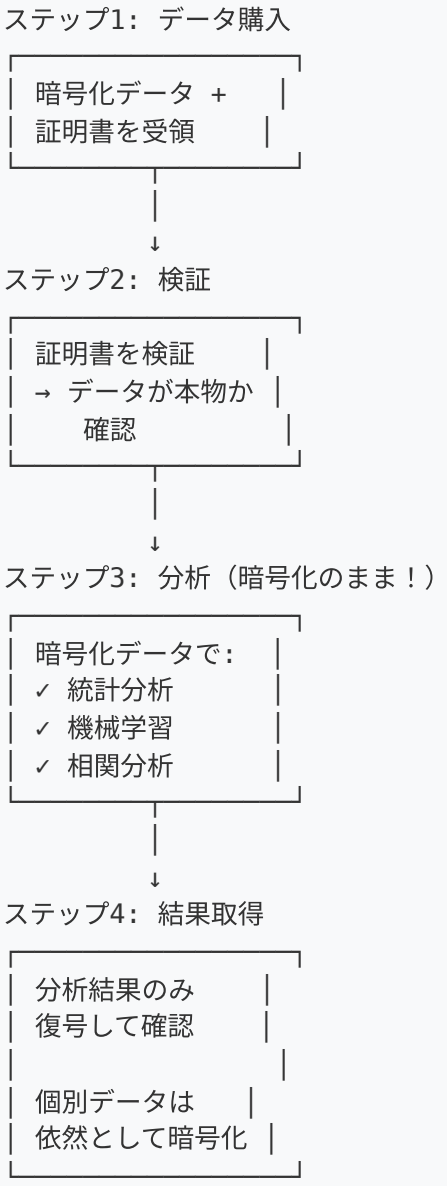
> データ提供者: 「このデータは本物で、改ざんされていません」
> データ購入者: 「データの中身を見ずに、それを確認したい」
>
> ZKP:
> - データの中身は見せない
> - でも、データが正しいことは数学的に証明
> - 改ざんは100%検出可能

システムの仕組み（簡単に）

データ提供者側（カルテ会社）のフロー



データ購入者側（製薬会社・研究機関）のフロー



☒ 重要なポイント

1. **データは常に暗号化**: 購入者も生データは見えない
2. **計算は可能**: 暗号化されたまま統計・機械学習ができる
3. **信頼性**: ZKPでデータの正当性が保証される

顧客が得られる具体的なメリット

データ提供者（カルテ会社・病院）のメリット

1. 完全なプライバシー保護

```
従来の仮名化:
患者データ → 仮名化 → 提供
                ↓
          再識別リスクあり

ZKP-DB:
患者データ → 暗号化 → 提供
                ↓
          再識別不可能（秘密鍵なしでは復号できない）
```

営業トーク:「万が一データが漏洩しても、暗号化されているので中身は絶対に見えません。患者のプライバシーは数学的に保護されます」

2. 法的リスクの低減

- GDPRやHIPAAに完全準拠
- 暗号化データは「個人データ」扱いされない（秘密鍵なしで復号不可能な場合）
- 訴訟リスクが大幅に低下

営業トーク:「法務部門の審査も通りやすくなります。GDPR対応も万全です」

3. 詳細データの提供が可能

```
従来: 集計データのみ提供
→ 「45歳以上の糖尿病患者の平均血圧」

ZKP-DB: 個別データを提供（暗号化済み）
→ 購入者が自由に分析できる
→ より高度な活用が可能
```

営業トーク:「より価値の高いデータ商品を提供できます。購入者の満足度が上がります」

4. 信頼性の証明

ZKPにより、データが本物であることを証明できる

> 購入者の不安:
> 「このデータ、本当に正確？」
> 「改ざんされていない？」
> 「古いデータじゃない？」
>
> ZKP-DBの回答:
> 「証明書で数学的に保証します」
> → 信頼性が大幅向上

データ購入者（製薬会社・研究機関）のメリット

1. より高度な分析が可能

> 従来の仮名化データ:
> ├─ 集計済みの統計データのみ
> ├─ カスタマイズした分析が困難
> └─ 機械学習の精度に限界
>
> ZKP-DB:
> ├─ 個別患者レベルのデータ（暗号化済み）
> ├─ 自由なクエリ実行
> ├─ 高精度な機械学習モデルを訓練可能
> └─ 詳細な相関分析

営業トーク:「暗号化されたままでも、御社が必要な分析を自由に実行できます。機械学習の精度も大幅に向上します」

2. データの信頼性を検証可能

> データ受領時:
> 1. 証明書を検証（数秒で完了）
> 2. データが改ざんされていないことを確認
> 3. データの品質を保証
>
> → 安心して使える

営業トーク:「データを受け取った瞬間に、そのデータが本物で改ざんされていないことを確認できます」

3. コンプライアンスリスクの低減

> 購入者の懸念:
> 「このデータを使って大丈夫？」
> 「プライバシー規制に抵触しない？」
> 「訴訟リスクは？」
>
> ZKP-DBの保証:
> ✓ データは完全に暗号化
> ✓ 個人の特定は不可能
> ✓ GDPRなどの規制に準拠
> ✓ 法的リスクが極めて低い

営業トーク:「御社の法務部門も安心できる、コンプライアンス完全対応のソリューションです」

4. 実用的な機械学習

> 可能なこと:
> ✓ 線形回帰モデル（完全対応）
> ✓ ロジスティック回帰（高精度）
> ✓ ニューラルネットワーク（2-3層）
> ✓ クラスタリング
> ✓ 相関分析
>
> 実用例:
> - 新薬の効果予測モデル
> - 副作用リスク評価
> - 患者層のセグメンテーション
> - 治療法の最適化

営業トーク:「実際のビジネスで必要な機械学習タスクは、ほぼすべて対応可能です」

セキュリティの保証

3層のセキュリティ

第1層: 準同型暗号

> セキュリティレベル: 軍事レベル
>
> 攻撃者が暗号化データを入手しても:
> × 中身を見ることができない
> × 秘密鍵なしでは復号不可能
> × スーパーコンピュータでも解読に数千年かかる
>
> 保護されるもの:
> ✓ 患者の年齢、性別
> ✓ 診断結果
> ✓ 治療履歴
> ✓ すべての個人情報

第2層: ゼロ知識証明

> 保証されること:
> ✓ データが改ざんされていない
> ✓ データが正しい範囲内にある
> ✓ データの出所が正当
>
> 検出できること:
> ✓ データの改ざん（100%検出）
> ✓ 不正なデータの混入
> ✓ 古いデータの使い回し

第3層: アクセス制御

> データ提供者側:
> ├─ 秘密鍵を厳重に管理
> ├─ 鍵の分散管理も可能
> └─ アクセスログを記録
>
> データ購入者側:
> ├─ 暗号化データのみアクセス可能
> ├─ 生データは見えない
> └─ 検証のみ実行可能

セキュリティの実証

シミュレーション: データ漏洩が発生した場合

シナリオ: サーバーがハッキングされ、データベースが盗まれた

> 仮名化システムの場合:
> 攻撃者が入手したデータ:
> 「45歳、男性、東京都、糖尿病、血圧145/90」
> ↓
> 他のデータと照合
> ↓
> 個人が特定される
> ↓
> × 重大なプライバシー侵害
>
> ZKP-DBの場合:
> 攻撃者が入手したデータ:
> 「0x7a4f9b2e3c8d1f6a9e2a7c5d...」(暗号文)
> ↓
> 復号を試みる
> ↓
> 秘密鍵がないので不可能
> ↓
> □ プライバシーは完全に保護される

営業トーク:「万が一の事態でも、患者のプライバシーは数学的に保護されます。これは保険のようなものです」

利用シーン別の説明

シーン1: 製薬会社の新薬開発

顧客のニーズ

「糖尿病の新薬を開発中。大規模な臨床データで効果を予測したい」

ZKP-DBのソリューション

```
ステップ1: データ購入
└ 10万人分の糖尿病患者データ（暗号化済み）

ステップ2: 検証
└ ZKPで正当性を確認（数秒）

ステップ3: 機械学習モデル訓練
└ 暗号化されたまま訓練
    - 年齢、性別、BMI、血糖値...
    - 治療法と効果の相関分析
    - 副作用リスクの予測

ステップ4: 予測モデル完成
└ 新薬の効果を高精度で予測
    - 開発期間の短縮
    - 成功確率の向上
```

メリット

- ☒ プライバシーを守りながら大規模データを活用
- ☒ 高精度な予測モデル
- ☒ コンプライアンスリスクなし

営業トーク:「患者のプライバシーを完全に守りながら、御社の新薬開発を加速できます」

シーン2: 保険会社のリスク評価

顧客のニーズ

「健康リスクを正確に評価して、適切な保険料を設定したい」

ZKP-DBのソリューション

```
ステップ1：データ購入
└ 年齢層別の健康データ（暗号化済み）

ステップ2：リスク分析
└ 暗号化されたまま分析
    - 年齢と疾患リスクの相関
    - ライフスタイルと健康の関係
    - 治療コストの予測

ステップ3：リスクモデル構築
└ より正確なリスク評価が可能に

ステップ4：保険商品の最適化
└ 適切な保険料設定
    - 顧客満足度向上
    - 収益性改善
```

メリット

- ☒ 大規模データで精度向上
- ☒ 規制当局への説明が容易
- ☒ 顧客への透明性を保てる

営業トーク:「プライバシー規制を完全にクリアしながら、より公平で精密なリスク評価が可能になります」

シーン3: 医療AI企業の診断支援システム開発

顧客のニーズ

「画像診断AIを開発したい。大量の症例データが必要」

ZKP-DBのソリューション

```
ステップ1: データ購入
└ 画像データ + 診断結果（暗号化済み）

ステップ2: AIモデル訓練
└ 暗号化されたまま訓練
    - ディープラーニングモデル
    - 診断精度の向上

ステップ3: モデル検証
└ 暗号化テストデータで評価

ステップ4: 実用化
└ 高精度な診断支援AIの完成
```

メリット

- ☒ 大規模データセットを安全に利用
- ☒ AIの精度向上
- ☒ 医療機関からの信頼獲得

営業トーク:「患者のプライバシーを守りつつ、最先端の診断AIを開発できます」

シーン4: 大学の医療研究

顧客のニーズ

「疫学研究で大規模なデータ分析をしたい」

ZKP-DBのソリューション

> ステップ1：データ取得
> └ 匿名化された大規模コホートデータ
>
> ステップ2：統計分析
> └ 暗号化されたまま分析
> - 疾患の発生率
> - リスク因子の特定
> - 治療効果の比較
>
> ステップ3：論文執筆
> └ 統計結果は取得可能
> - 個人データは一切不要
> - 倫理委員会の承認が容易
>
> ステップ4：研究成果の公開
> └ プライバシーを守った研究

メリット

- ☒ 倫理審査がスムーズ
- ☒ 大規模データで研究の質向上
- ☒ 国際的な共同研究も可能

営業トーク:「倫理的な問題をクリアしながら、世界最高水準の医療研究が可能です」

よくある質問と回答

Q1: 本当に暗号化されたまま計算できるのですか？

A: はい、可能です。これを「準同型暗号」と呼びます。

分かりやすい例:

> 封筒に入った2つの数字を、封筒を開けずに足し算する
>
> 通常：封筒を開けないと計算できない
> 準同型暗号：封筒を開けずに計算できる魔法のような技術

実用例:

> 暗号化された血圧データ100人分
> ↓
> 暗号化されたまま平均を計算
> ↓
> 結果のみ復号
> ↓
> 「平均血圧: 128 mmHg」
>
> ※個別の患者データは一切見ていない

Q2: どんな計算ができますか？

A: 多くの統計・機械学習タスクに対応しています。

完全対応:

- ☒ 平均、分散などの統計量
- ☒ 線形回帰
- ☒ 相関分析
- ☒ データの集計

高精度対応:

- ☒ ロジスティック回帰（分類）
- ☒ ニューラルネットワーク（2-3層）
- ☒ クラスタリング

制限あり:

- ⚠ 深いニューラルネットワーク（5層以上）
- ⚠ Random ForestやXGBoost

解決策: 複雑なモデルが必要な場合は、ハイブリッド方式や対話型計算などの高度な手法を使用できます（別途説明）

Q3: 処理速度はどのくらいですか？

A: 通常の計算より遅いですが、実用レベルです。

```
100人のデータで平均を計算:
通常の計算: 0.001秒
暗号化計算: 0.1秒（100倍）

10,000人のデータで回帰分析:
通常の計算: 1秒
暗号化計算: 数分

※ただし技術は急速に進化中
```

重要なポイント:

- データ分析は「一度実行すれば良い」場合が多い
- 多少時間がかかっても、プライバシー保護のメリットが大きい
- バッチ処理で夜間実行も可能

Q4: ゼロ知識証明とは何ですか？

A: 「情報を明かさずに、その情報が正しいことを証明する技術」です。

日常の例:

```
あなたが20歳以上であることを証明したい

通常の方法:
運転免許証を見せる
→ 年齢だけでなく、住所や顔写真も見られてしまう

ゼロ知識証明:
「20歳以上である」ことだけ証明
→ 具体的な年齢や他の情報は一切明かさない
```

医療データへの応用:

```
データ提供者: 「このデータは本物です」
証明書: 「確かに本物だと数学的に保証します」

購入者が確認できること:
✓ データが改ざんされていない
✓ データが正しい範囲内にある
✓ データの出所が正当

購入者が見えないもの:
× データの中身
× 個別の患者情報
```

Q5: 従来の仮名化とどう違いますか？

A: セキュリティレベルが根本的に異なります。

```
┌──────────────┬──────────────┬──────────────┐
│    項目      │    仮名化    │    ZKP-DB    │
├──────────────┼──────────────┼──────────────┤
│データ形式    │ 平文         │ 暗号文       │
│再識別リスク  │ あり（87%）  │ なし（0%）   │
│漏洩時の影響  │ 重大         │ 軽微         │
│データ突合    │ 可能         │ 不可能       │
│法的リスク    │ 中〜高       │ 低           │
│提供可能範囲  │ 集計データ   │ 詳細データ   │
└──────────────┴──────────────┴──────────────┘
```

詳細は別資料「仮名化vs暗号技術」を参照

Q6: コストはどのくらいかかりますか？

A: 初期投資は必要ですが、長期的には投資対効果が高いです。

計算機資源:

- サーバー: 通常より高性能なものが必要
- ストレージ: 暗号化データは元データの2-3倍
- 計算時間: 通常の10-100倍

ただし:

- クラウドサービスの活用で初期投資を抑制可能
- バッチ処理で効率化
- 技術の進化で年々高速化

Q7: 既存のシステムと統合できますか？

A: はい、API経由で統合可能です。

```
既存システム（病院の電子カルテ等）
    ↓ API連携
ZKP-DBシステム
    ↓
暗号化 + 証明書生成
    ↓
データマーケットプレイス
```

統合方法:

1. REST API: 標準的なHTTP通信
2. バッチ処理: 定期的にデータをエクスポート
3. リアルタイム連携: ストリーミングデータにも対応可能

Q8: 将来的に技術が古くなりませんか？

A: 最新の暗号技術を採用しており、将来性があります。

技術の進化:

- 準同型暗号: 2009年に理論確立、現在実用化段階
- ゼロ知識証明: 暗号通貨（ブロックチェーン）で広く採用
- 両技術とも急速に発展中

将来性:

- 各国政府・大企業が研究開発に投資
- 標準技術になる可能性が高い
- 計算速度は年々向上（Mooreの法則）

アップグレード:

- 暗号アルゴリズムは切り替え可能
- システムの設計が柔軟

用語集

- **準同型暗号**: 暗号化されたまま計算できる暗号技術
- **ゼロ知識証明**: 情報を明かさずに正しさを証明する技術
- **仮名化**: 個人を特定できる情報を削除・置換する方法
- **GDPR**: 欧州のデータ保護規則
- **HIPAA**: 米国の医療情報保護法

まとめ: 本システムのメリット

顧客に伝えるべき3つのメッセージ

1. **完全なプライバシー保護** 「数学的に保証された、最高レベルのプライバシー保護です」

2. **実用的な機能** 「暗号化されたままでも、必要な分析・機械学習が可能です」
3. **信頼性の保証** 「ゼロ知識証明により、データの正当性を証明できます」

NGワード（避けるべき表現）

- ×「絶対に安全」→ ☒ 「数学的に保護されている」
- ×「すべての計算が可能」→ ☒ 「多くの実用的な計算に対応」
- ×「速度は変わらない」→ ☒ 「実用レベルの速度」

自信を持って言えること

- ☒ 秘密鍵なしでは復号不可能
- ☒ 再識別リスクはゼロ
- ☒ 改ざんは100%検出可能
- ☒ GDPR等の規制に準拠
- ☒ 統計分析・機械学習が可能