「外部に出せない『大切なデータ』を、
自らの手で資産に変える挑戦の記録」

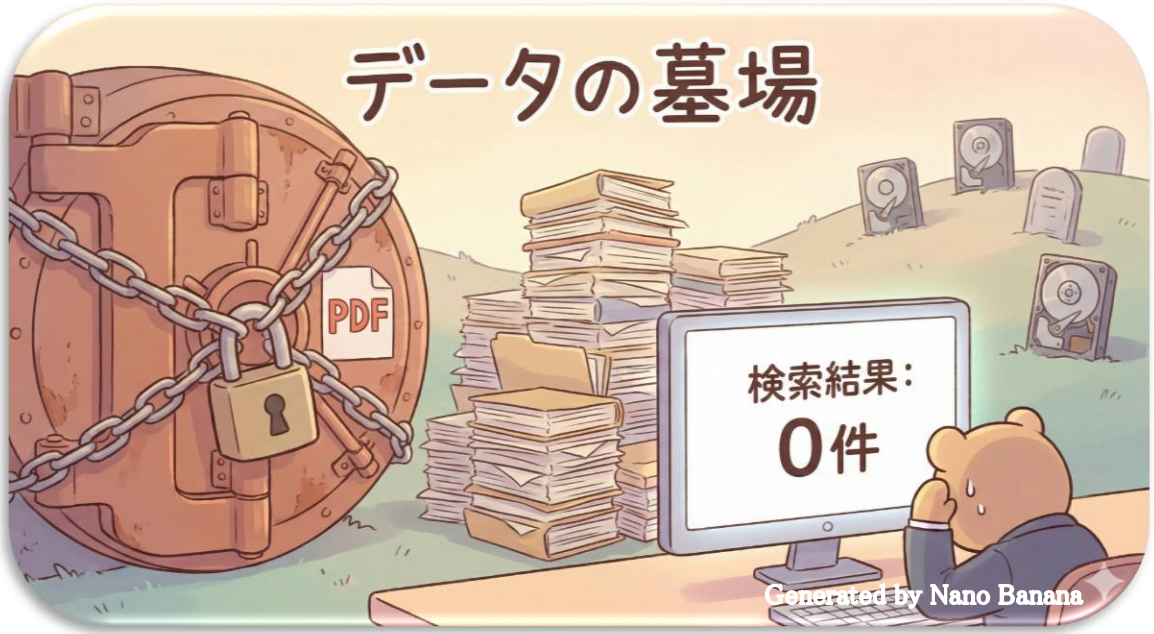
マルチモーダルLLMのローカル運用のススメ

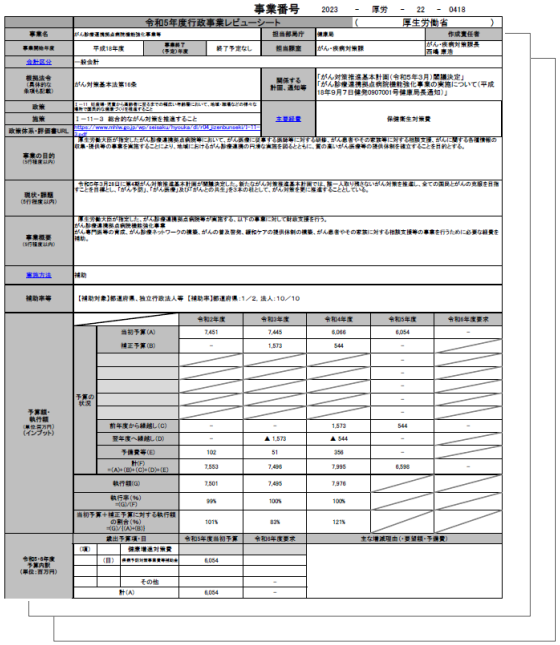
～手元に眠る機微PDFを「お宝」に変えるデータ整備インフラ～

2026年1月10日

データデザイナー　柴田修一

背景：PDFという名の「開かない金庫」





事業番号　2023 - 厚労 - 22 - 0418

事業名	令和5年度行政事業レビューシート			厚生労働省		
予算の状況						
当初予算(A)	7,451	7,445	6,066	6,054		
補正予算(B)	-	1,573	544	-		
前年度から繰越し(C)	-	-	1,572	544		
翌年度へ繰越し(D)	-	▲ 1,573	▲ 544	-		
予備費等(E)	102	51	356	-		
計(F) =(A)+(B)+(C)+(D)+(E)	7,553	7,496	7,995	6,598		
執行額(G)	7,501	7,495	7,976			
執行率(%) =(G)/(F)	99%	100%	100%			
当初予算+補正予算に対する執行額の割合(%) =(G)/(A)+(B)	101%	82%	121%			

令和5年度予算内訳（単位：百万円）	歳出予算項・目	令和5年度当初予算	令和5年度要求	主な増減理由（※要望額・予備費）
(項)	健康増進対策費			
(目)	疾病予防対策事業等補助金	6,054		
	その他	-		
	計(A)	6,054	-	

複雑な構造を有する
「行政事業レビューシート」

現場に眠る「死蔵データ」の現実：

医療、介護、経営の最前線。

貴重な記録ほどPDFという「開かない金庫」に閉じ込められ、検索も分析もできないまま「墓場」に眠っています。

「死蔵」から「資産」への転換：

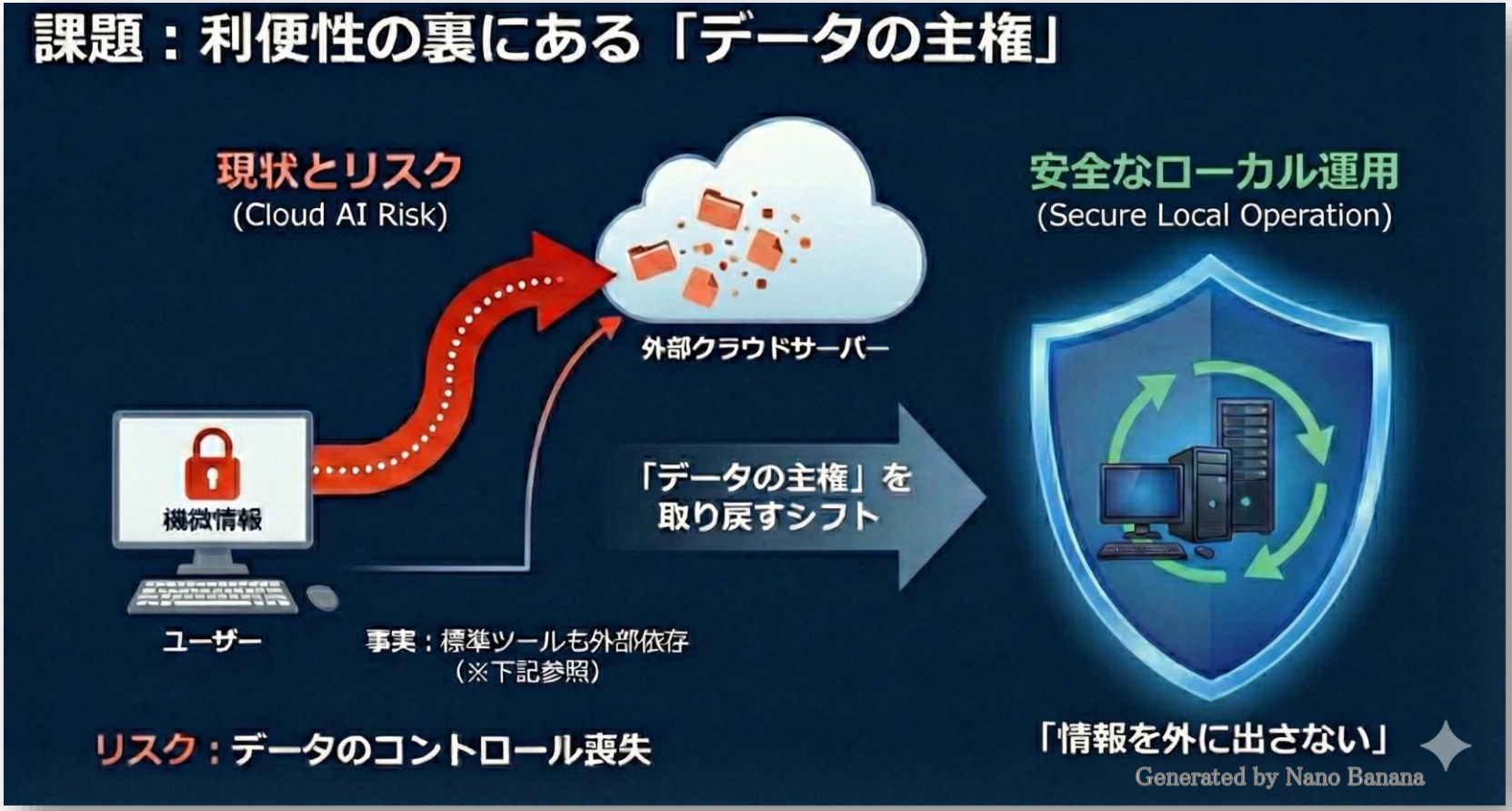
もし、これらの情報を検索・活用できる形式へ救い出せたら？

それは組織の意思決定を支える「分析資産」への第一歩となります。

【実証テスト】データ化の壁にどこまで迫れるか：

あえて「行政事業レビューシート」という、屈指の複雑さを誇るドキュメントを技術検証用のサンプルに選び、ローカルAIの抽出能力をテストしてみました。

機密データを「クラウド」に渡せるか？

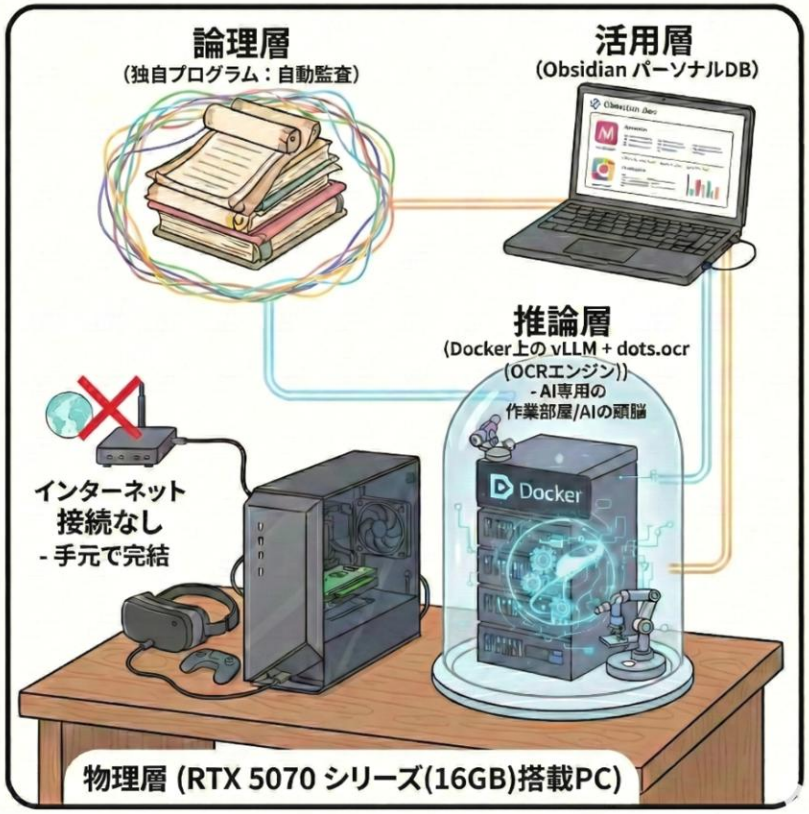


PDF処理で代表的なA社の現行ツールは、
デスクトップ、Web、モバイルが統合されたソリューションと定義されており、
Webベースの処理（クラウド利用）がその一部に含まれている。

「公式FAQにおいて、『デスクトップ、モバイル、web にまたがって統合されたソリューション』であり、
『アプリケーションとサービスが含まれる』と明記されています。
Webやサービスが統合されている以上、利用方法によってはクラウド上でデータが処理される構成になっています。」
https://helpx.adobe.com/jp/document-cloud/faq.html

システム構成：手元で完結する「データ工廠（こうしょう）」

ポイント：「インターネットへの出口を断つ」ことで得られる圧倒的な安心感。



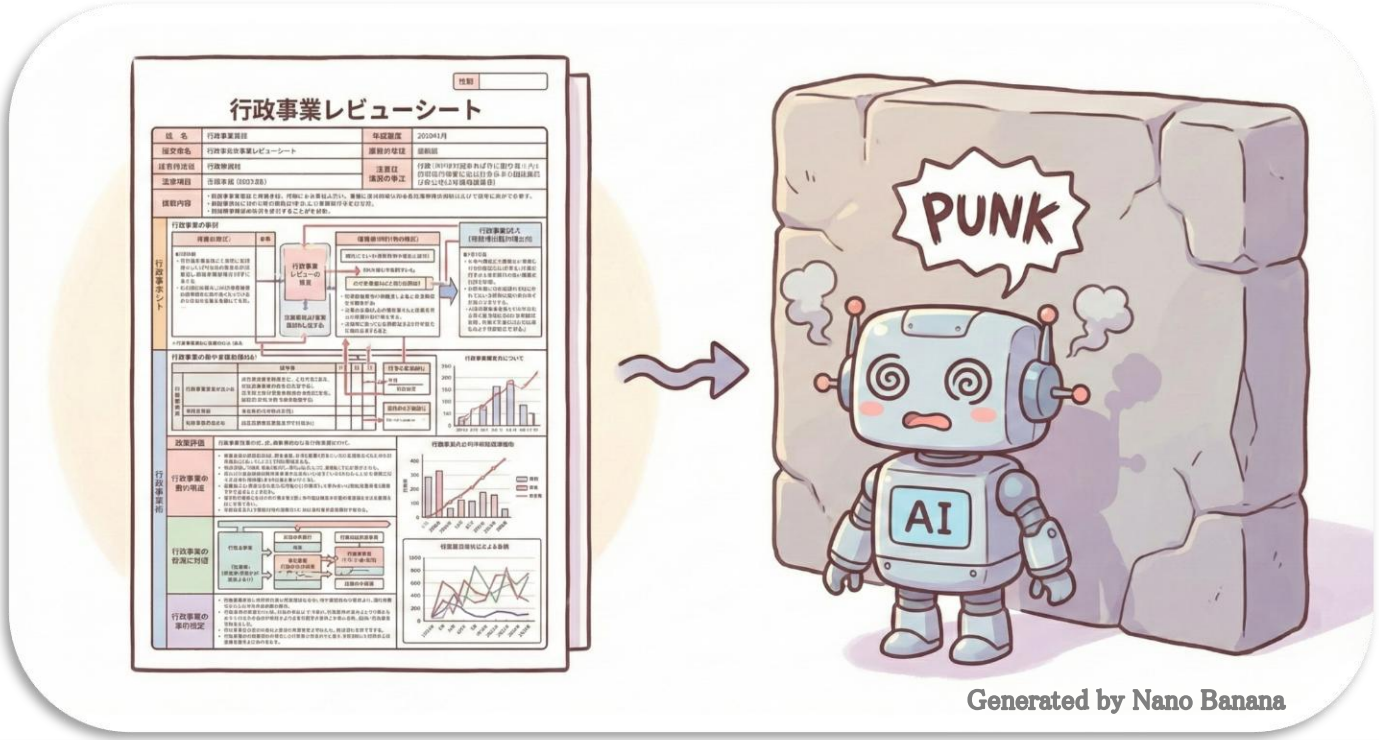
「情報を守る」環境作り：
外部送信を伴わないため、
機微情報の漏洩リスクを最小限に抑えられます。
組織内での安全なデータ活用のための
現実的な選択肢です。

「専用のリソース」で着実に：
PCの計算力をAI処理に割り振り、
日常業務を妨げることなく、
一歩ずつ確実にデータ化を進めることができます。

「多層チェック」で精度を補う：
AI単体では不完全な読み取りを、
独自プログラムが自動監査。
不備を洗い出し、
人間が確認・修正しやすい形に整えます。

「自由な形式」で蓄積：
特定のサービスや有償ソフトに依存せず、
将来にわたって編集・再加工がしやすい
データ形式で保存します。

実証テスト：難攻不落の「行政事業レビューシート」に挑む



検証用サンプル：厚労省「がん・疾病対策」関連資料

非常に複雑で高密度な表構造を持つ資料を、
ローカルAIの限界と実力を測るための「高難度の実証テスト用サンプル」として採用しました。

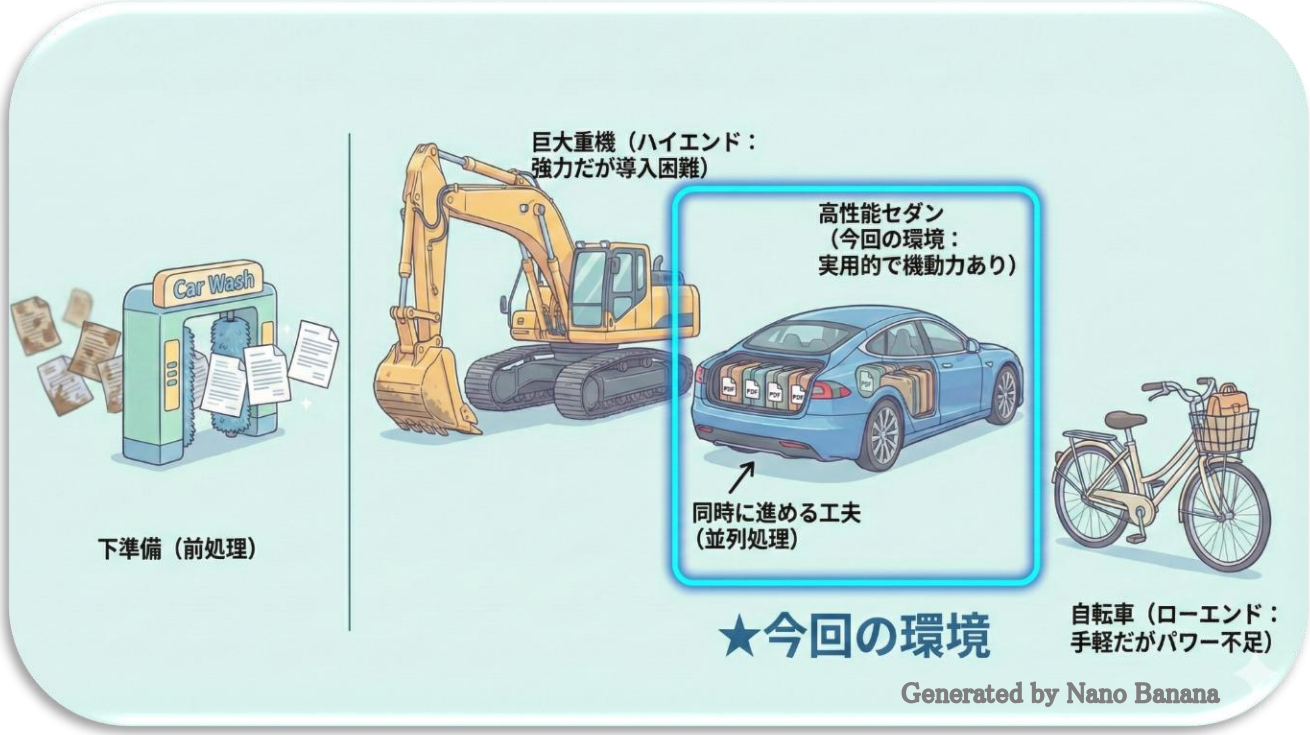
テストの意義：複雑な非定型ドキュメントの攻略

このレベルの構造をデータ化する土台ができれば、
世の中にある一般的な業務PDFの多くで、活用のハードルが大きく下がります。

直面した「物理的な壁」：AIの視界の限界

高精細な情報を1枚丸ごと処理しようとすると、AIの処理能力を超えてしまい、
情報の省略や欠落が発生するという限界に突き当たりました。

チューニング：手の届きやすい「実用的な性能」で効率を最大化する



ハードウェア：一般に普及している「高性能な実用モデル」を採用

専門機関向けの超高性能機ではなく、入手しやすい範囲でAI処理に高い適性を持つ「高性能な実用モデル（RTX 5070 シリーズ）」を使用しています。

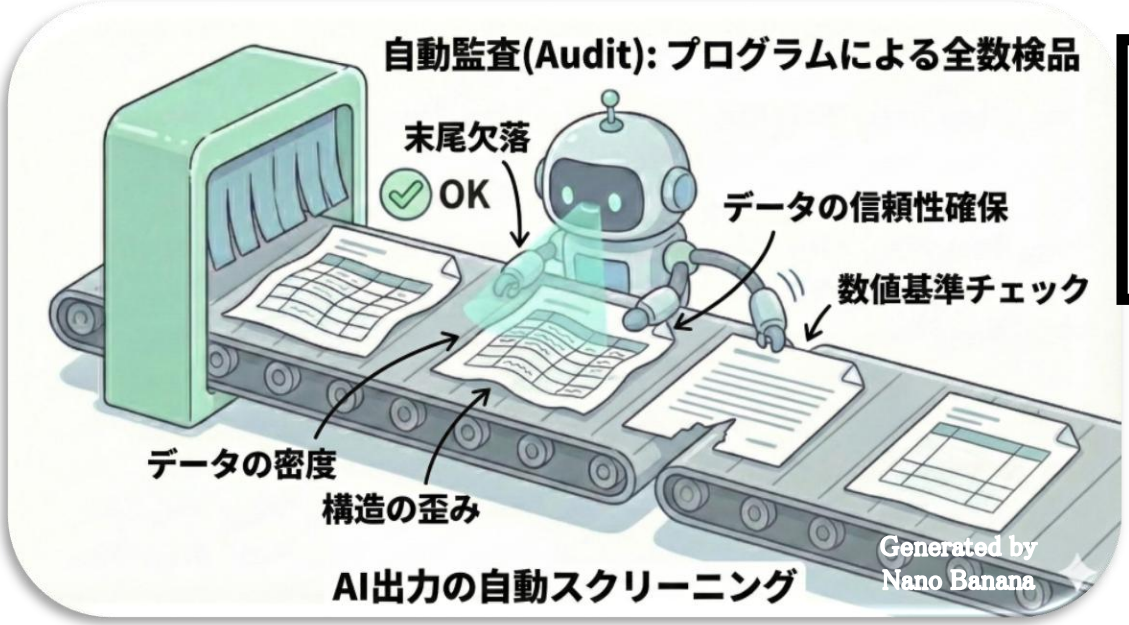
同時に進める工夫：限られた作業スペース（16GBメモリ）を使い切る

1枚ずつの速さを追うのではなく、AIが一度に使える「作業用のメモリ（16GB）」をフル活用。
全体で「いかに多くの枚数を同時にこなすか」という効率を優先しました。

下準備（前処理）：AIが読みやすい状態へプログラムで一括変換

画像の鮮明化などを独自のスクリプトで自動処理。
AIが最も得意とする「視界がクリアな状態」でデータを供給し、精度の質を底上げします。

監査の要：AI の「不完全さ」を自動で検知する




```
--- [Host] Running OCR Quality Audit ---
--- [Host] Starting Dynamic Integrity Audit ---
✅ [OK] r05_rv03a_day1-001.json   | Dens:119.0 | Ratio:0.41 | EOF:True
✅ [OK] r05_rv03a_day1-002.json   | Dens:162.2 | Ratio:0.47 | EOF:True
❌ [NG] r05_rv03a_day1-003.json   | Dens: 58.6 | Ratio:0.29 | EOF:True  | ERR: THIN_DATA(58.6c/r)
❌ [NG] r05_rv03a_day1-004.json   | Dens:  0.0 | Ratio:0.00 | EOF:True  | ERR: EMPTY_OR_INVALID
✅ [OK] r05_rv03a_day1-005.json   | Dens:135.2 | Ratio:0.35 | EOF:True
✅ [OK] r05_rv03a_day1-006.json   | Dens:161.1 | Ratio:0.37 | EOF:True
------------------------------------------------------------
Audit Complete. Found 2 files requiring revival.
------------------------------------------------------------
```

自動監査 (Audit)：プログラムによる「全数検品」

AI が途中で力尽きていないか（末尾欠落）、内容を勝手に省略していないか（データの密度）、
表がガタガタになっていないか（構造の歪み）を、独自のスクリプトが厳しい数値基準でチェックします。

検知の重要性：不備のあるデータを可能な限り資産に混ぜない

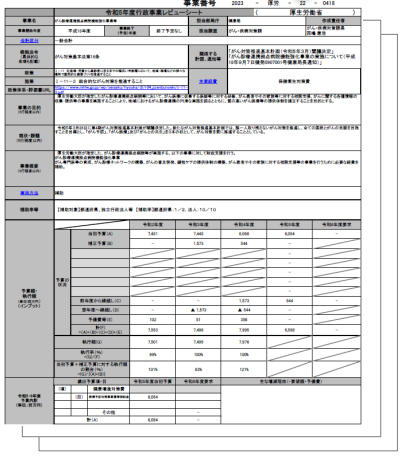
全てのエラーを完璧に見抜けるわけではありませんが、
AIの出力を鵜呑みにせず自動でスクリーニングを行うこと。
この「ブレーキ」を組み込むことが、
ローカル運用におけるデータの信頼性を可能な限り高めるための不可欠なプロセスとなります。

【検証用サンプル】厚生労働省 行政事業レビューシート

公開プロセス（令和5年度）

行政事業レビュー（公開プロセス）の実施について

開催日	対象事業	資料等							動画
2023/6/2	未払賃金立替払事務実施費	レビューシート [273KB]	事業概要・見直し案 [1,079KB]	論点等 [108KB]	全体版 [17,905KB]	コメント結果 [529KB]	外部有識者のコメント [142KB]	議事録 [329KB]	Youtube
	求職者支援制度に必要な経費	レビューシート [813KB] セグメントシート [209KB]	事業概要・見直し案 [2,375KB]	論点等 [100KB]			外部有識者のコメント [150KB]	議事録 [341KB]	Youtube
	がん診療連携拠点病院機能強化事業等	レビューシート [...KB]	事業概要・見直し案 [1,349KB]	論点等 [112KB]			外部有識者のコメント [134KB]	議事録 [318KB]	Youtube
	在宅福祉事業費補助金	ビューシート [...14KB]	事業概要・見直し案 [2,118KB]	論点等 [99KB]			外部有識者のコメント [123KB]	議事録 [313KB]	Youtube
	歯科衛生士に対する復職支援・離職防止等推進事業	レビューシート [555KB]	事業概要・見直し案 [1,547KB]	論点等 [103KB]			外部有識者のコメント [138KB]	議事録 [321KB]	Youtube
	社会福祉施設等施設整備費補助金	レビューシート [311KB]	事業概要・見直し案 [6,167KB]	論点等 [65KB]			外部有識者のコメント [131KB]	議事録 [329KB]	Youtube



複雑な構造を有する
「行政事業レビューシート」

出所：「行政事業レビュー 公開プロセス（令和5年度）」（厚生労働省）

データの出所：
令和5年（2023年）6月2日に実施された「公開プロセス」の公開資料より選定。

ピックアップした事業：
当日取り上げられた6つの事業の中から、
「がん診療連携拠点病院機能強化事業等」を精度検証用の素材として採用。

ドキュメントの規模：
全6ページにわたる高密度な表構造。

実証のポイント：
極めて高度な専門性と、複雑に入り組んだ表構造を持つドキュメントを、
汎用AIがいかに構造化できるか、その「データ化の可能性」を探ります。

【実例1】成功：構造の再現に成功した例

C

	支出先	法人番号	業務概要	支出額（百万円）	契約方式等	入札者数（応募者数）	落札率	一者応札・一者応募又は競争性のない随意契約となった理由及び改善策（支出額10億円以上）
1	国立研究開発法人国立がん研究センター	6010005015219	がんゲノム情報管理センター事業の実施等	2,880	補助金等交付	-	-	-
2	国立研究開発法人国立成育医療研究センター	6010905002126	がん診療連携拠点病院機能強化事業の実施	122	補助金等交付	-	-	-
3	独立行政法人国立病院機構九州がんセンター	1013205001281	がん診療連携拠点病院機能強化事業の実施	101	補助金等交付	-	-	-
4	国立大学法人東北大学病院	7370005002147	がん診療連携拠点病院機能強化事業の実施	99	補助金等交付	-	-	-
5	国立大学法人京都大学医学部附属病院	3130005005532	がん診療連携拠点病院機能強化事業の実施	88	補助金等交付	-	-	-
6	国立大学法人九州大学病院	3290005003743	がん診療連携拠点病院機能強化事業の実施	87	補助金等交付	-	-	-
7	国立大学法人三重大学医学部附属病院	2190005003044	がん診療連携拠点病院機能強化事業の実施	87	補助金等交付	-	-	-
8	国立大学法人東海国立大学機構名古屋大学医学部附属病院	3180005006071	がん診療連携拠点病院機能強化事業の実施	86	補助金等交付	-	-	-
9	国立大学法人広島大学広島大学病院	1240005004054	がん診療連携拠点病院機能強化事業の実施	72	補助金等交付	-	-	-
10	独立行政法人国立病院機構四国がんセンター	1013205001281	がん診療連携拠点病院機能強化事業の実施	71	補助金等交付	-	-	-

D

	支出先	法人番号	業務概要	支出額（百万円）	契約方式等	入札者数（応募者数）	落札率	一者応札・一者応募又は競争性のない随意契約となった理由及び改善策（支出額10億円以上）
1	富士通株式会社	1020001071491	がんゲノム情報レポジトリーシステム更改 一式	558	一般競争契約（総合評価）	1	84.9%	-
2	株式会社日立製作所	7010001008844	がんゲノム医療情報利活用システム構築	440	随意契約（その他）	-	100%	-
3	三井情報株式会社	6010401078785	ゲノム医療知識統合システム構築	364	随意契約（その他）	-	100%	-
4	富士通株式会社	1020001071491	がんゲノム情報レポジトリシステム構築	270	随意契約（その他）	-	100%	-
5	富士通株式会社	1020001071491	ハードウェア・ソフトウェア保守、データセンター・回線利用	269	随意契約（その他）	-	100%	-
6	三井情報株式会社	6010401078785	運用支援業務	182	一般競争契約（最低価格）	1	99%	-
7	株式会社日立製作所	7010001008844	運用支援業務	111	一般競争契約（最低価格）	1	99%	-
8	富士通株式会社	1020001071491	ヘルプデスク業務	110	一般競争契約（最低価格）	1	99%	-
9	クラスメソッド株式会社	5011101037603	クラウド利用料	69	一般競争契約（最低価格）	1	100%	-
10	富士通株式会社	1020001071491	運用支援業務	26	一般競争契約（最低価格）	1	99%	-

支出先上位10者リスト欄についてさらに記載が必要な場合はチェックの上【別紙3】に記載　チェック

抽出データ

C

	支出先	法人番号	業務概要	支出額（百万円）	契約方式等	入札者数（応募者数）	落札率	一者応札・一者応募又は競争性のない随意契約となった理由及び改善策（支出額10億円以上）
1	国立研究開発法人国立がん研究センター	6010005015219	がんゲノム情報管理センター事業の実施等	2,880	補助金等交付	-	-	-
2	国立研究開発法人国立成育医療研究センター	6010905002126	がん診療連携拠点病院機能強化事業の実施	122	補助金等交付	-	-	-
3	独立行政法人国立病院機構九州がんセンター	1013205001281	がん診療連携拠点病院機能強化事業の実施	101	補助金等交付	-	-	-
4	国立大学法人東北大学病院	7370005002147	がん診療連携拠点病院機能強化事業の実施	99	補助金等交付	-	-	-
5	国立大学法人京都大学医学部附属病院	3130005005532	がん診療連携拠点病院機能強化事業の実施	88	補助金等交付	-	-	-
6	国立大学法人九州大学病院	3290005003743	がん診療連携拠点病院機能強化事業の実施	87	補助金等交付	-	-	-
7	国立大学法人三重大学医学部附属病院	2190005003044	がん診療連携拠点病院機能強化事業の実施	87	補助金等交付	-	-	-
8	国立大学法人東海国立大学機構名古屋大学医学部附属病院	3180005006071	がん診療連携拠点病院機能強化事業の実施	86	補助金等交付	-	-	-
9	国立大学法人広島大学広島大学病院	1240005004054	がん診療連携拠点病院機能強化事業の実施	72	補助金等交付	-	-	-
10	独立行政法人国立病院機構四国がんセンター	1013205001281	がん診療連携拠点病院機能強化事業の実施	71	補助金等交付	-	-	-

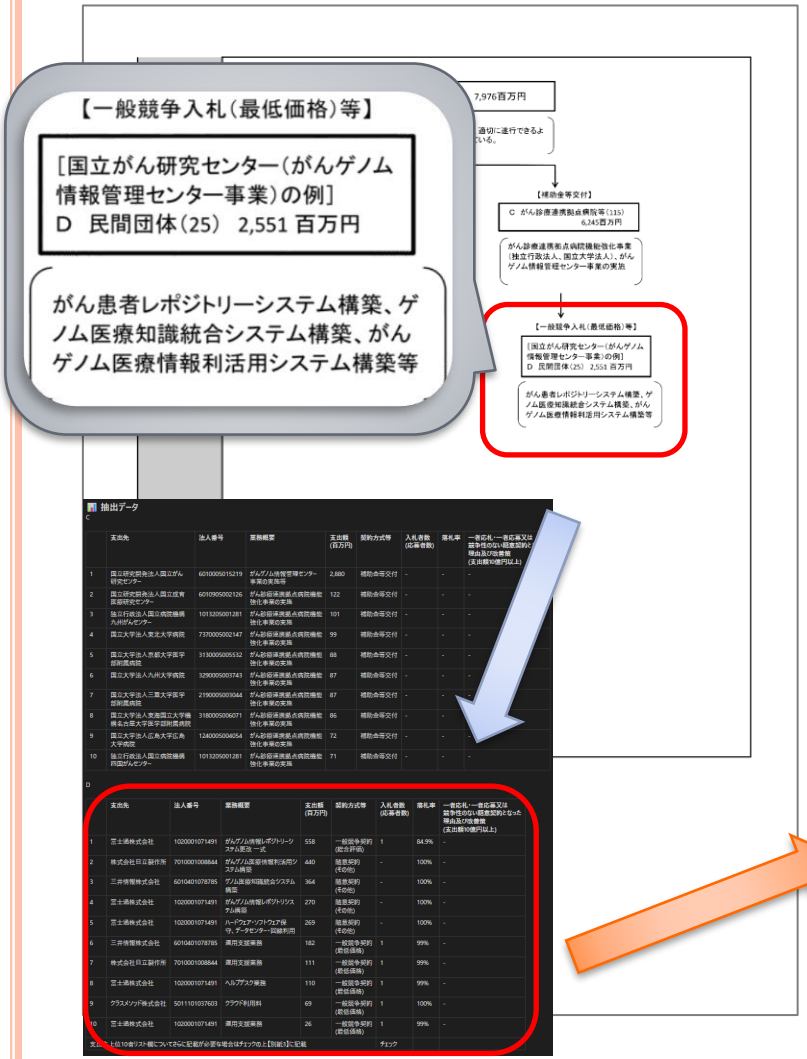
D

	支出先	法人番号	業務概要	支出額（百万円）	契約方式等	入札者数（応募者数）	落札率	一者応札・一者応募又は競争性のない随意契約となった理由及び改善策（支出額10億円以上）
1	富士通株式会社	1020001071491	がんゲノム情報レポジトリーシステム更改 一式	558	一般競争契約（総合評価）	1	84.9%	-
2	株式会社日立製作所	7010001008844	がんゲノム医療情報利活用システム構築	440	随意契約（その他）	-	100%	-
3	三井情報株式会社	6010401078785	ゲノム医療知識統合システム構築	364	随意契約（その他）	-	100%	-
4	富士通株式会社	1020001071491	がんゲノム情報レポジトリシステム構築	270	随意契約（その他）	-	100%	-
5	富士通株式会社	1020001071491	ハードウェア・ソフトウェア保守、データセンター・回線利用	269	随意契約（その他）	-	100%	-
6	三井情報株式会社	6010401078785	運用支援業務	182	一般競争契約（最低価格）	1	99%	-
7	株式会社日立製作所	7010001008844	運用支援業務	111	一般競争契約（最低価格）	1	99%	-
8	富士通株式会社	1020001071491	ヘルプデスク業務	110	一般競争契約（最低価格）	1	99%	-
9	クラスメソッド株式会社	5011101037603	クラウド利用料	69	一般競争契約（最低価格）	1	100%	-
10	富士通株式会社	1020001071491	運用支援業務	26	一般競争契約（最低価格）	1	99%	-

別紙3】に記載　チェック

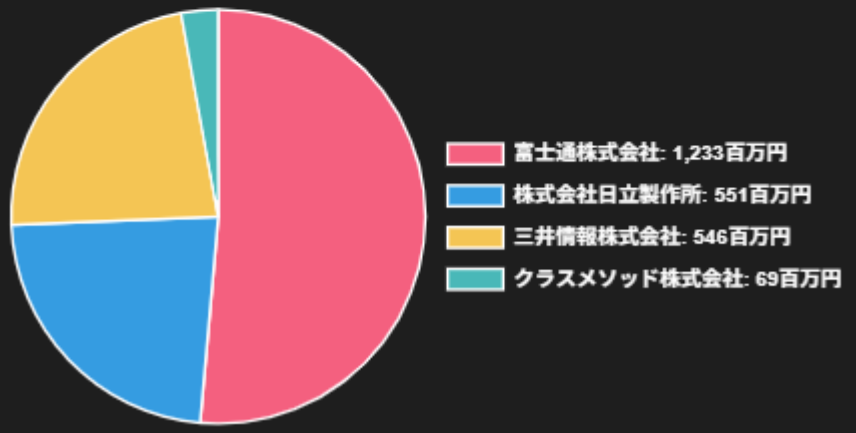
"fileName": "r05_rv03a_day1-001",
"text": "事業番号
2023 - 厚労 - 22 - 0418
<table><thead><tr><th colspan=¥"5¥">令和5年度行政事業レビューシート</th><th>(厚生労働省)</th></tr></tr><tr><th>事業名</th><td colspan=¥"3¥">がん診療連携拠点病院機能強化事業等</td></tr><tr><th>担当部局庁</th><td>健康局</td><th>作成責任者</th></tr><tr><td>平成18年度</td><th>事業終了
(予定)年度</th><td>終了予定なし</td><th>担当課室</th><td>がん・疾病対策課</td><td>がん・疾病対策課長
西嶋 康浩</td></tr><tr><th>会計区分</th><td colspan=¥"4¥">一般会計</td><td></td><td></td><td></td><td></td></tr><tr><th>根拠法令
(具体的な
条項も記載)</th><td colspan=¥"3¥">がん対策基本法第16条</td><th>関係する
計画、通知等</th><td colspan=¥"2¥">「がん対策推進基本計画(令和5年3月)閣議決定」
「がん診療連携拠点病院機能強化事業の実施について(平成
18年9月7日健発0907001号健康局長通知)」</td></tr><tr><th>政策</th><td colspan=¥"3¥">I-11 妊産婦・小児から高齢者に至るまでの幅広い年齢層において、地域・職場などの様々な場所で国民的な健康づくりを推進すること</td><th rowspan=¥"2¥">主要経費</th><td rowspan=¥"2¥">保健衛生対策費</td></tr><tr><th>施策</th><td colspan=¥"3¥">I-11-3 総合的ながん対策を推進すること

成果：PDFの奥底から掘り出した「知見」



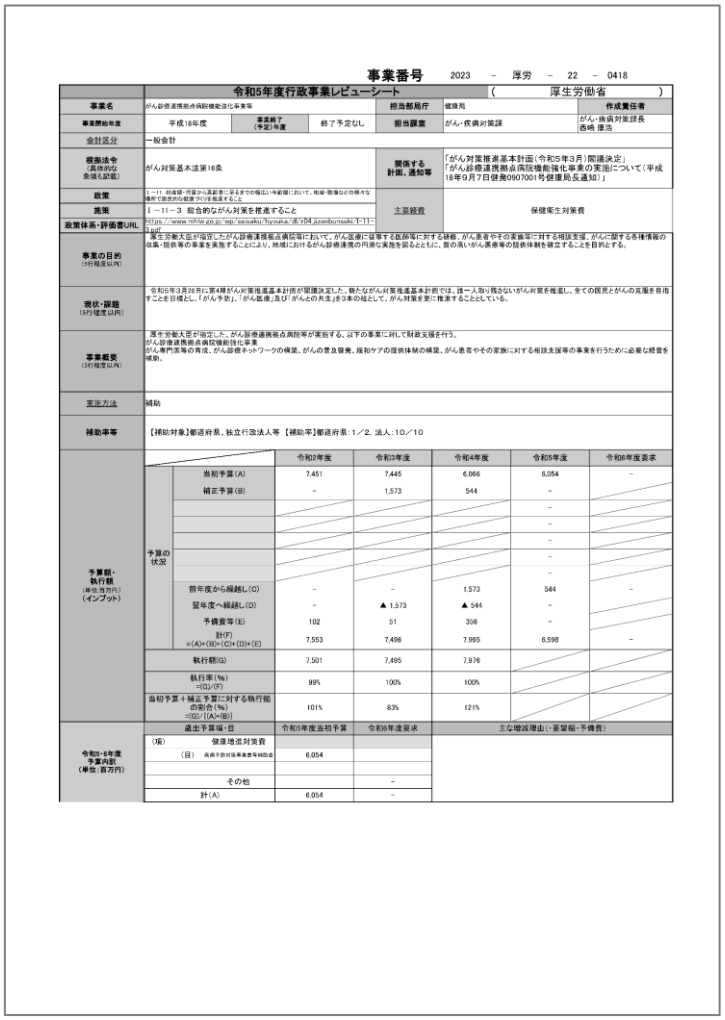
がんゲノム 情報管理センター事業 契約上位10者（2022執行分）

支出先 (10)	事業概要	金額（百万円）
富士通株式会社	がんゲノム情報レポジトリーシステム更改 一式	558
株式会社日立製作所	がんゲノム医療情報利活用システム構築	440
三井情報株式会社	ゲノム医療知識統合システム構築	364
富士通株式会社	がんゲノム情報レポジトリシステム構築	270
富士通株式会社	ハードウェア・ソフトウェア保守、データセンター・回線利用	269
三井情報株式会社	運用支援業務	182
株式会社日立製作所	運用支援業務	111
富士通株式会社	ヘルプデスク業務	110
クラスメソッド株式会社	クラウド利用料	69
富士通株式会社	運用支援業務	26



10者合計: 2,399 百万円

【実例2】課題：密度と複雑さでズレが生じた例



事業番号　2023 － 厚労 － 22 － 0418

令和5年度行政事業レビューシート　（厚生労働省）

事業名	がん診療連携拠点病院機能強化事業等		担当部局庁	健康局	作成責任者
事業開始年度	平成18年度	事業終了（予定）年度	終了予定なし	担当課室	がん・疾病対策課長 西嶋 康浩
会計区分	一般会計				
根拠法令（具体的な条項も記載）	がん対策基本法第16条		関係する計画、通知等	「がん対策推進基本計画（令和5年3月）閣議決定」 「がん診療連携拠点病院機能強化事業の実施について（平成18年9月7日健発0907001号健康局長通知）」	
政策	Ⅰ－11 妊産婦・児童から高齢者に至るまでの幅広い年齢層において、地域・職場などの様々な場所で国民的な健康づくりを推進すること		主要経費	保健衛生対策費	
施策	Ⅰ－11－3　総合的ながん対策を推進すること				
政策体系・評価書URL	https://www.mhlw.go.jp/wp/seisaku/hyouka/dl/r04_jizenbunseki/I-11-3.pdf				

事業の目的（5行程度以内）	厚生労働大臣が指定したがん診療連携拠点病院等において、がん医療に従事する医師等に対する研修、がん患者やその家族等に対する相談支援、がんに関する各種情報の収集・提供等の事業を実施することにより、地域におけるがん診療連携の円滑な実施を図るとともに、質の高いがん医療等の提供体制を確立することを目的とする。
現状・課題（5行程度以内）	令和5年3月28日に第4期がん対策推進基本計画が閣議決定した。新たながん対策推進基本計画では、誰一人取り残さないがん対策を推進し、全ての国民とがんの克服を目指すことを目標とし、「がん予防」、「がん医療」及び「がんとの共生」を3本の柱として、がん対策を更に推進することとしている。
事業概要（5行程度以内）	厚生労働大臣が指定した、がん診療連携拠点病院等が実施する、以下の事業に対して財政支援を行う。 がん診療連携拠点病院機能強化事業 がん専門医等の育成、がん診療ネットワークの構築、がんの普及啓発、緩和ケアの提供体制の構築、がん患者やその家族に対する相談支援等の事業を行うために必要な経費を補助。
実施方法	補助
補助率等	【補助対象】都道府県、独立行政法人等【補助率】都道府県：1／2，法人：10／10

予算額・執行額（単位：百万円）（インプット）		令和2年度	令和3年度	令和4年度	令和5年度	令和6年度要求
予算の状況	当初予算(A)	7,451	7,445	6,066	6,054	－
	補正予算(B)	－	1,573	544	－	
	前年度から繰越し(C)	－	－	1,573	544	－
	翌年度へ繰越し(D)	－	▲ 1,573	▲ 544	－	
	予備費等(E)	102	51	356	－	
	計(F) =(A)+(B)+(C)+(D)+(E)	7,553	7,496	7,995	6,598	－
執行額(G)		7,501	7,495	7,976		
執行率(%) =(G)/(F)		99%	100%	100%		
当初予算＋補正予算に対する執行額の割合(%) =(G)/[(A)+(B)]		101%	83%	121%		

令和5・6年度予算内訳（単位：百万円）		歳出予算項・目	令和5年度当初予算	令和6年度要求	主な増減理由（・要望額・予備費）
	(項)	健康増進対策費			
	(目)	疾病予防対策事業費等補助金	6,054		
		その他		－	
		計(A)	6,054	－	

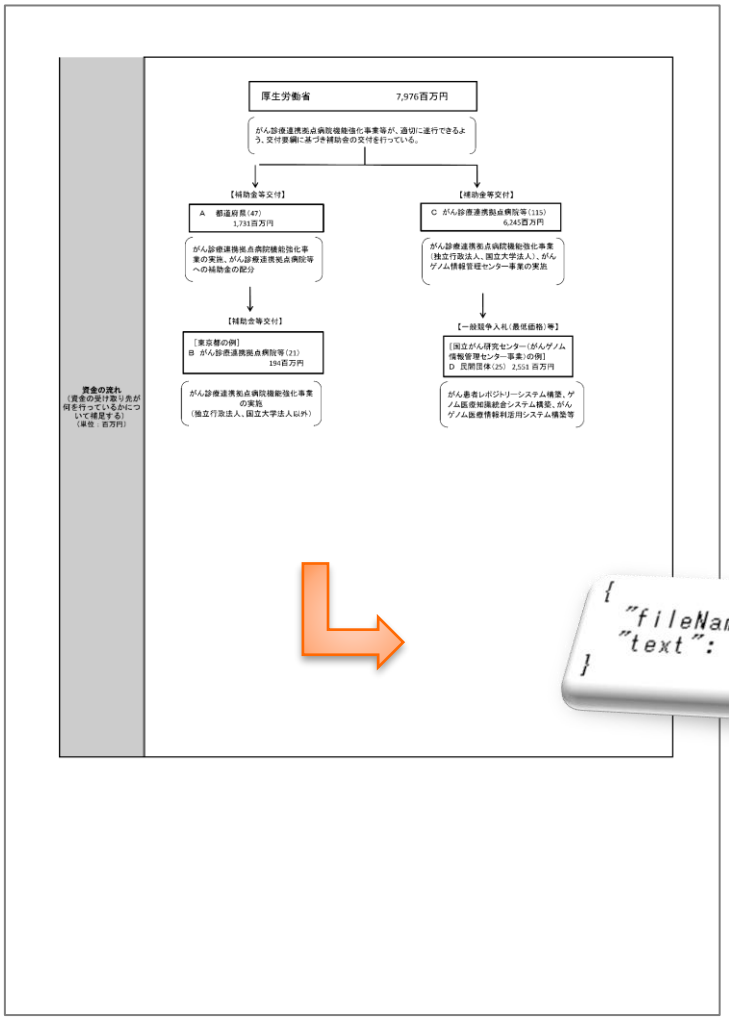
抽出データ

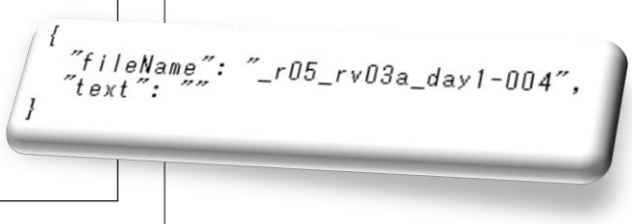
事業番号

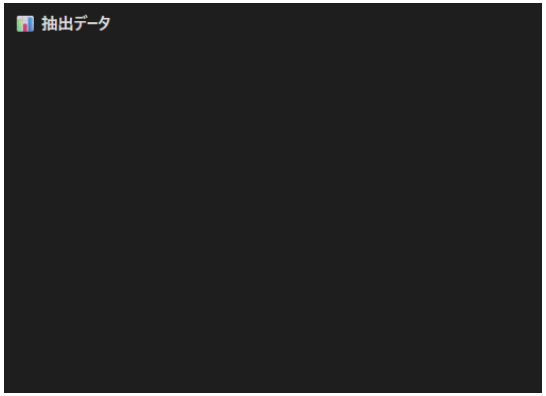
2023 - 厚労 - 22 - 0418

令和5年度行政事業レビューシート					(厚生労働省)		
事業名	がん診療連携拠点病院機能強化事業等			担当部局庁	健康局	作成責任者	
事業開始年度	平成18年度	事業終了(予定)年度	終了予定なし	担当課室	がん・疾病対策課	がん・疾病対策課長 西嶋 康浩	
会計区分	一般会計						
根拠法令(具体的な条項も記載)	がん対策基本法第16条			関係する計画、通知等	「がん対策推進基本計画(令和5年3月)閣議決定」 「がん診療連携拠点病院機能強化事業の実施について(平成18年9月7日健発0907001号健康局長通知)」		
政策	I-11 妊産婦・児童から高齢者に至るまでの幅広い年齢層において、地域・職場などの様々な場所で国民的な健康づくりを推進すること			主要経費	保健衛生対策費		
施策	I-11-3 総合的ながん対策を推進すること https://www.mhlw.go.jp/wp/seisaku/hyouka/dl/r04_jizenbunseki/I-11-3.pdf						
政策体系・評価書URL							
事業の目的(5行程度以内)	厚生労働大臣が指定したがん診療連携拠点病院等において、がん医療に従事する医師等に対する研修、がん患者やその家族等に対する相談支援、がんに関する各種情報の収集・提供等の事業を実施することにより、地域におけるがん診療連携の円滑な実施を図るとともに、質の高いがん医療等の提供体制を確立することを目的とする。						
現状・課題(5行程度以内)	令和5年3月28日に第4期がん対策推進基本計画が閣議決定した。新たながん対策推進基本計画では、誰一人取り残さないがん対策を推進し、全ての国民とがんの克服を目指すことを目標とし、「がん予防」、「がん医療」及び「がんとの共生」を3本の柱として、がん対策を更に推進することとしている。						
事業概要(5行程度以内)	厚生労働大臣が指定した、がん診療連携拠点病院等が実施する、以下の事業に対して財政支援を行う。 がん診療連携拠点病院機能強化事業 がん専門医等の育成、がん診療ネットワークの構築、がんの普及啓発、緩和ケアの提供体制の構築、がん患者やその家族に対する相談支援等の事業を行うために必要な経費を補助。						
実施方法	補助						
補助率等	【補助対象】都道府県、独立行政法人等【補助率】都道府県:1/2,法人:10/10						
予算額・執行額(単位:百万円)(インプット)			令和2年度	令和3年度	令和4年度	令和5年度	令和6年度要求
予算の状況	当初予算(A)			7,451	7,445	6,066	6,054	-
	補正予算(B)			-	1,573	544	-	-
	前年度から繰越し(C)			-	-	1,573	544	-
	翌年度へ繰越し(D)			-	▲ 1,573	▲ 544	-	-
	予備費等(E)			102	51	356	-	-
	計(F) =(A)+(B)+(C)+(D)+(E)			7,553	7,496	7,995	6,598	-
	執行額(G)			7,501	7,495	7,976	-	-
	執行率(%) =(G)/(F)			99%	100%	100%	-	-
	当初予算+補正予算に対する執行額の割合(%) =(G)/[(A)+(B)]			101%	83%	121%	-	-
歳出予算項・目			令和5年度当初予算	令和6年度要求	主な増減理由(・要望額・予備費)			
令和5・6年度予算内訳(単位:百万円)	(項)	健康増進対策費						
	(目)	疾病予防対策事業費等補助金	6,054					
		その他		-				
		計(A)	6,054	-				

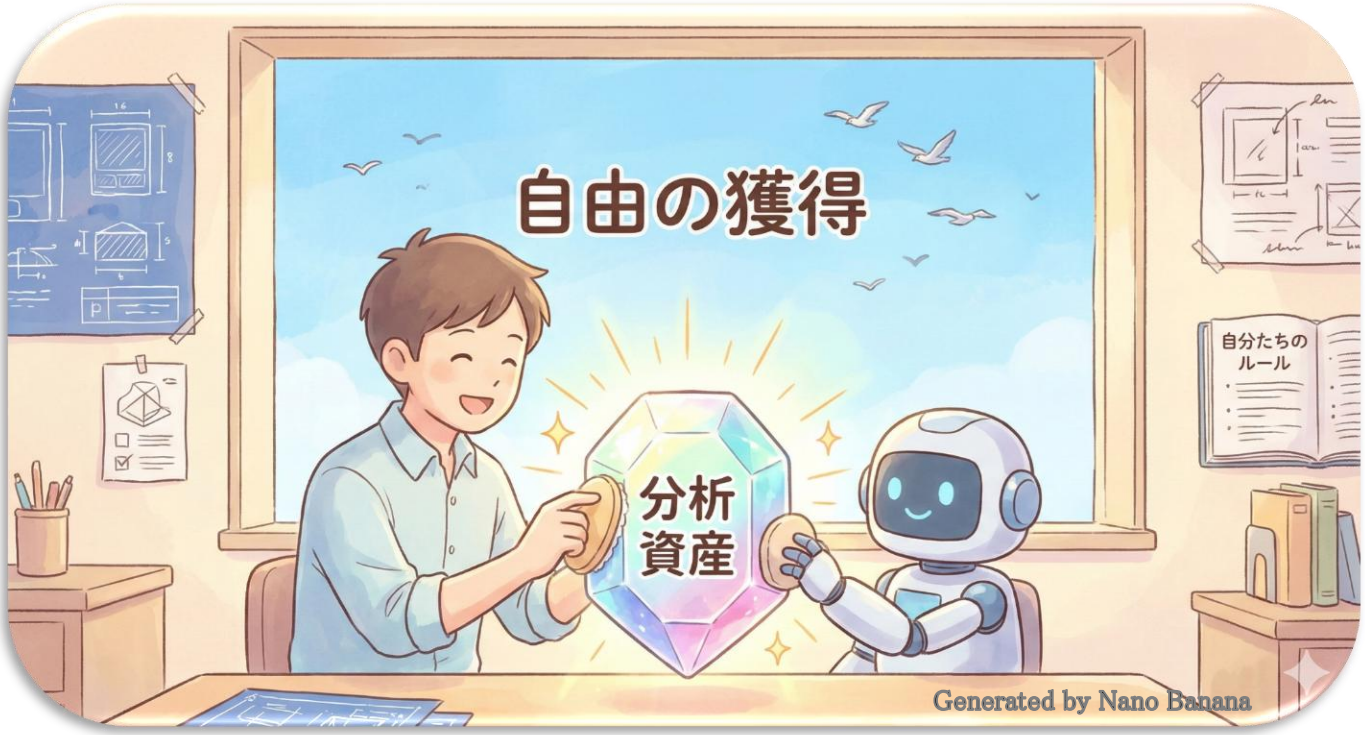
【実例3】例外：チャート構造で抽出対象外となった例







考察:ローカル LLM が実現する「データ主権」



活用を支える安全な基盤:
機密情報を一歩も外に出さず、現場で自在にハンドリングできる環境。

自律的な資産化サイクル:
外部プラットフォームの制約に縛られず、組織の知見を独自のルールで磨き上げる仕組み。

制約からの解放と創造性:
コストや利用規約を気にせず、知見を自在に操る「自由」の獲得。

ローカル運用は、**データの主権**を取り戻し、
組織の可能性を自ら広げる**「自由」の獲得**に繋がるものと考えます。

将来展望：専門環境と巨大モデルが切り拓く可能性

【実例2】と同じページをより大きな計算機リソースを使った結果

抽出データ

Table 1

ステータス

継続: 独立 ー 完結 | 列数: 7

	令和5年度行政事業レビューシート			(厚生労働省)		
事業名	がん診療連携拠点病院機能強化事業等			担当部局庁	健康局	作成責任者
事業開始年度	平成18年度	事業終了(予定)年度	終了予定なし	担当課室	がん・疾病対策課	がん・疾病対策課長 西嶋康浩
会計区分	一般会計					
根拠法令(具体的な条項も記載)	がん対策基本法第16条					
政策	Ⅰ－11 妊産婦・児童から高齢者に至るまでの幅広い年齢層において、地域・職場などの様々な場所で国民的な					
施策	Ⅰ－11－3 総合的ながん対策を推進すること					
政策体系・評価書URL	https://www.mhlw.go.jp/wp/seisaku/hyouka/dl/r04_jizenbunseki/Ⅰ－11－3.pdf					
事業の目的(5行程度以内)	厚生労働大臣が指定したがん診療連携拠点病院等において、がん医療に従事する医師等に対する研修、がん患者やその家族等に対する相談支援、がんに関する各種情報の収集・提供等の事業を実施することにより、地域におけるがん診療連携の円滑な実施を図るとともに、質の高いがん医療等の提供体制を確立することを目的とする。					
現状・課題(5行程度以内)	令和5年3月28日に第4期がん対策推進基本計画が閣議決定した。新たながん対策推進基本計画では、誰一人取り残さないがん対策を推進し、全ての国民とがんの克服を目指すことを目標とし、「がん予防」「がん医療」及び「がんとの共生」を3本の柱として、がん対策を更に推進することとしている。					
事業概要(5行程度以内)	厚生労働大臣が指定した、がん診療連携拠点病院等が実施する、以下の事業に対して財政支援を行う。がん診療連携拠点病院機能強化事業 がん専門医等の育成、がん診療ネットワークの構築、がんの普及啓発、緩和ケアの提供体制の構築、がん患者やその家族に対する相談支援等の事業を行うために必要な経費の一部。					
実施方法	補助					
補助率等	[補助対象]都道府県、独立行政法人等 [					

予算額・執行額(単位:百万円)(インプット)	令和2年度	令和3年度	令和4年度	令和5年度	令和6年度要求	
	当初予算(A) 補正予算(B)	7,451	7,445	6,066	6,054	-
	予算の状況	-	1,573	544	-	
	前年度から繰越し(C)	-	-	1,573	544	-
	翌年度へ繰越し(D)	-	▲1,573	▲544	-	
	予備費等(E)	102	51	356	-	
	計(F)=(A)+(B)+(C)+(D)+(E)	7,553	7,496	7,995	6,598	-
	執行額(G)	7,501	7,495	7,976		
	執行率(%):=(G)/(F)	99%	100%	100%		
令和5・6年度						
計(A)						

年度表記行におけるカラム位置の不整合（列ズレ）

進化と残存する課題：
上位モデル採用により構造再現性は向上。
一方で、年度行など極めて複雑なセル結合の認識には、依然として課題が残存。

信頼性向上の見通し：
モデル側の認識能力の向上や入力の工夫により、こうした課題が段階的に改善され、実用的な信頼性がより高まっていくことが期待される。

計算リソースとモデルの最適化により、
データ化の信頼性はさらに向上していくものと考えています。

結び：守り、救い、活かす

- **本発表のまとめ：**
  「機密を守り、墓場からデータを救い出し、意思決定に活かす」。
  このローカル LLM インフラは、あらゆる現場の死蔵データを資産に変える力を持っていると考えます。

- **さらなる展開：**
  今回の「紙（PDF）のデータ化」は、以前公開した「音声のデータ化（ローカル要約）」に続く第2弾の取り組みです。

- **ブログ記事紹介：**
  音声のローカル活用については、ブログ記事
  「NVIDIA Blackwell (GB10) 上で vLLM と Whisper を共存させるレシピ」
  にて詳しく公開しています。
  ※Yahoo検索「NVIDIA Whisper」にて検索結果 5位（約300万件中）にランクインしている、技術記事です。

泥臭い工夫の先にある、あなただけの強力な武器となるデータをその手に。

ご清聴ありがとうございました。