

知財DXでイノベーション時間を拡大

# THE調査力 *AI*



アイ・ピー・ファイン株式会社

# THE調査力 *AI*

(Deskbee)

# AI機能活用の手引き

2024年11月



改訂1(10/20)  
改訂2(11/01)  
改訂3(11/16)  
改訂4(12/03)  
改訂5(01/06)  
改訂6(02/01)  
改訂7(22/05)  
改訂8(22/10)  
改訂9(23/11)  
改訂10(24/11)

# アイ・ピー・ファインのAI機能

当社のAI機能は、GoogleのTensorFlowをエンジンとしています。

当社のAI機能は、6年間の実績と知財AI活用研究会活動を通して、AI活用のノウハウの修得に努めています。そして、そのポイントは次の2点に集約されます。

## 使い方の重要ポイント

### ① AI判定結果の品質は、「教師特許」の選択で決まる

(注釈)

2017年2月15日にTensorFlow 1.0がリリースされ、このオープンソース公開の理由は、**AIの真の価値はAIの「エンジン」ではなく、AIを賢くするのに必要な「データ」である。** Googleがオープンソースとして公開した理由をこう述べています。

### ② 判定精度と効率化は、類似順で分割する「二値分類」の方法により決まる

# お薦めポイント5つ

- ① サーチとノイズの2教師型AIで、関連特許とノイズ特許の分離を行います。
- ② 社内で評価済の特許を教師とすることで、AIがこれを基準として判定処理します。

## お薦めポイント5つ

- ① FTO調査、動向調査、先行文献調査のいずれも活用できます
- ② スタートは、SDI調査のノイズ判定に活用しノウハウを修得します
- ③ 海外特許調査は、Japio-GPG/FXの日本語利用がお勧めです
- ④ 遡及調査は、一定期間(1年程度)の人手評価により教師を作ります
- ⑤ AIによる自動判定で過去評価済特許の見直しも可能になります

# AI処理結果値とその活用

AI処理で出力されるデータは、“サーチ確率値”と“ノイズ確率値”の2つのみです。THE調査力AIにより両確率差を算出し昇順に特許の並べ替えを行い判定順位を付与します。この順位こそが重要な「値」となります。そして、判定順位の上位から何%までをサーチ判定に、以降をノイズ判定とします。これは次ページ「標準境界値の決め方」を参照ください。

## AI処理後、確率差構成比で境界値を設定

THE調査力AI | スクリーニング・AIナビゲータ theme 000005 : リチウムイオン二次電池に関する日本特許調査【サンプル】

タイトル一覧 AI判定履歴

AIによる類似特許判定結果の確認と、キーワードを使った教師特許の再設定による再学習が行えます。

確率差構成比で設定する場合

(※)整数を入力してください。  
 (※)サーチ判定とノイズ判定は、合計が100%以内になるようにしてください。

サーチ判定  % ノイズ判定  % ← 調査全体のノイズ比率×70%を初期値として推奨

サーチ判定は、確率差が上位からの設定となります。  
 ノイズ判定は、確率差が下位からの設定となります。

判定結果の採用： 確率差構成比  スクリーニング・AIナビゲータ

判定順位が大事

判定結果

	A	B	AI処理値	F	G
1	判定順位		確率差 (1)-(2)	社内評価	出願番号
2	1	サーチ特許と判定されたもの	0.3	ゴミ	特願2019-220018
3	2	サーチ特許と判定されたもの	-3.8	ゴミ	特願2019-216984
4	3	サーチ特許と判定されたもの	-4.9	ゴミ	特願2019-214414
5	4	サーチ特許と判定されたもの	-7.3	ゴミ	特願2020-213530
6	5	サーチ特許と判定されたもの	-8.7	ゴミ	特願2019-214755
7	6	サーチ特許と判定されたもの	-13.4	ゴミ	特願2019-213008
8	7	サーチ特許と判定されたもの	-13.8	ゴミ	特願2020-208012
9	8	サーチ特許と判定されたもの	-14.5	△	特願2019-214816
10	9	サーチ特許と判定されたもの	-14.6	ゴミ	特願2019-218275
11	10	サーチ特許と判定されたもの	-15.2	△	特願2019-225476
78	77	ノイズ特許と判定されたもの	-42.9	ゴミ	特願2020-566325
79	78	ノイズ特許と判定されたもの	-43.1	ゴミ	特願2019-212846
80	79	ノイズ特許と判定されたもの	-43.1	ゴミ	特願2019-215975
81	80	ノイズ特許と判定されたもの	-43.3	ゴミ	特願2020-547146
82	81	ノイズ特許と判定されたもの	-43.4	ゴミ	特願2019-214546
83	82	ノイズ特許と判定されたもの	-43.8	ゴミ	特願2020-523654

確率差でソートした「順位」が大事

サーチ判定

ノイズ判定

# 標準境界値の決め方

当社のAI機能で「二値分類」のために大事なポイントは、何を基準としてサーチ判定とノイズ判定に分類するかということです。ここで言う「標準境界値」は初期値として推奨するもので、検証結果やSDIの継続利用の経験などを基に上位側(効率UP)／下位側(精度UP)に変更が可能です。

## 標準境界値算出の解説

一定期間(例えば1年間)の調査件数全件に占めるノイズ特許の比率から算出します。概ね、90%がノイズ特許である場合、この70%をノイズ判定とする場合、全体の63%がノイズ判定値となり、逆に37%がサーチ判定値となります。この両値が初期値として推奨している「標準境界値」です。

例. A評価の特許:30件、ノイズ評価の特許:270件



# AI活用によるリスクの考え方

## AI活用の基本

**AI処理** 文書の類似度を数値化(順位付けと二値分類)

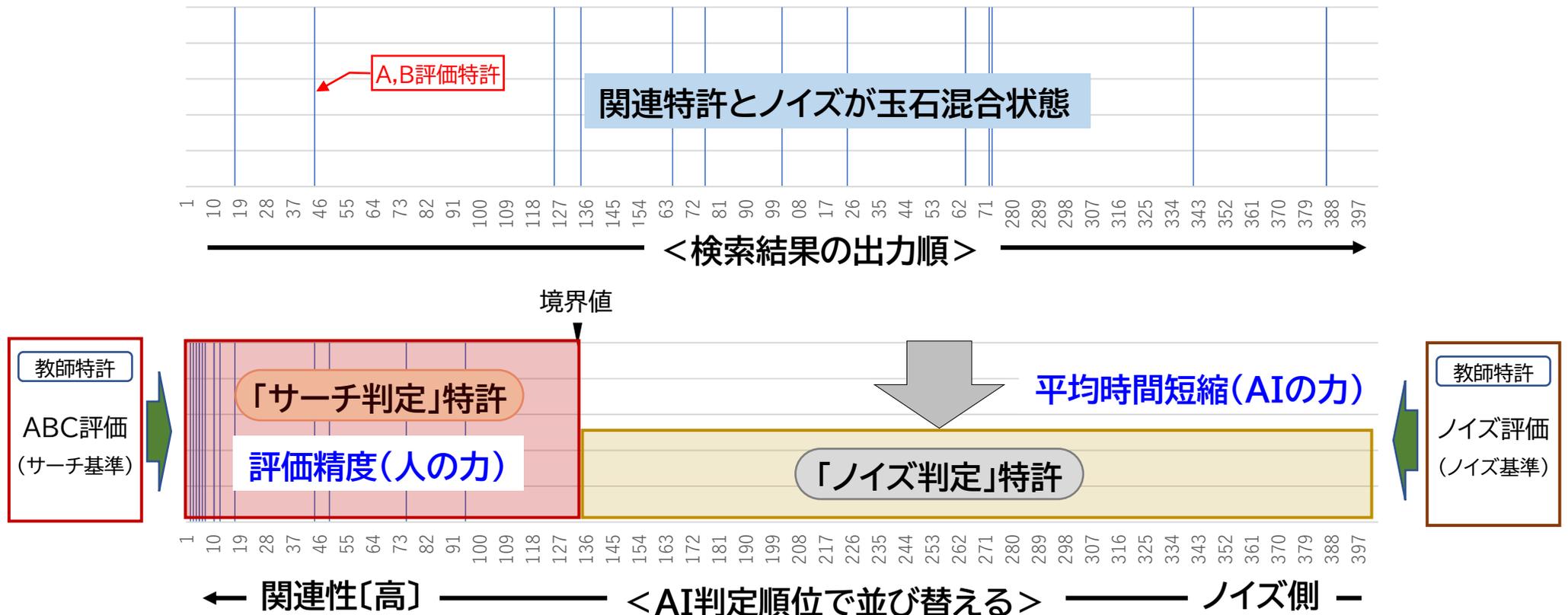


**運用** 閲覧・評価の時間に濃淡がつけられる

- ◆ AI活用の基本を間違わなければ、効率化を達成しつつリスクはこれまでの調査と変わりません

現 状

AI活用

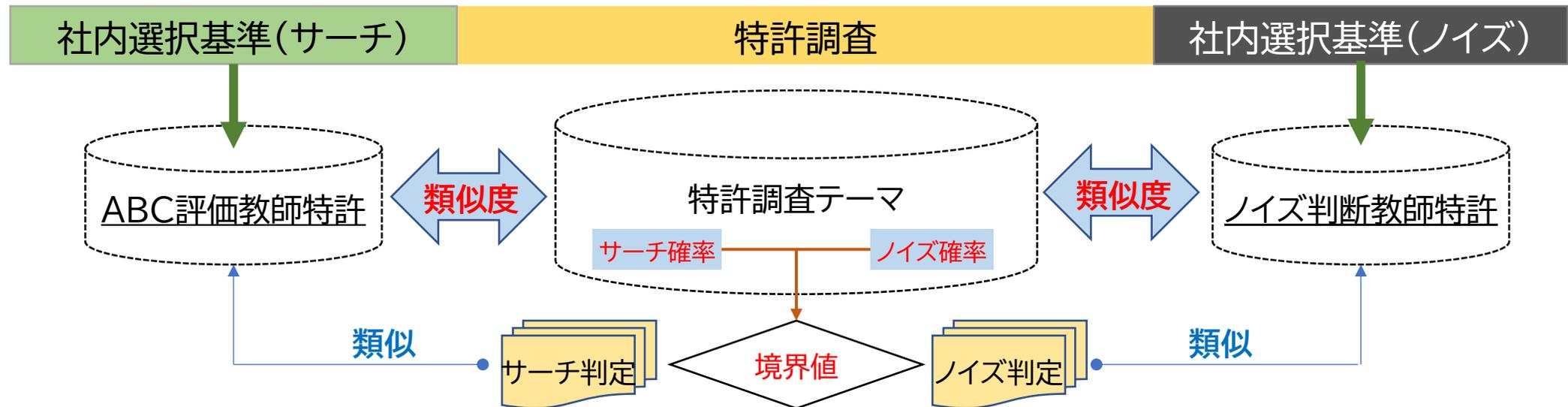


# 教師特許で「社内選択基準」を教える

AI機能は、テキストデータのみを使って処理を行います。従って、利用者が何を必要としどれが不要なのかを判断できる情報がありません。

そこで教師特許を準備して、関連特許とノイズ特許を教師として入力することにより、それぞれの教師データと判定される文書の類似度を数値化し出力します。

特許調査における教師とは、特許調査における「社内選択基準」であるということになります。

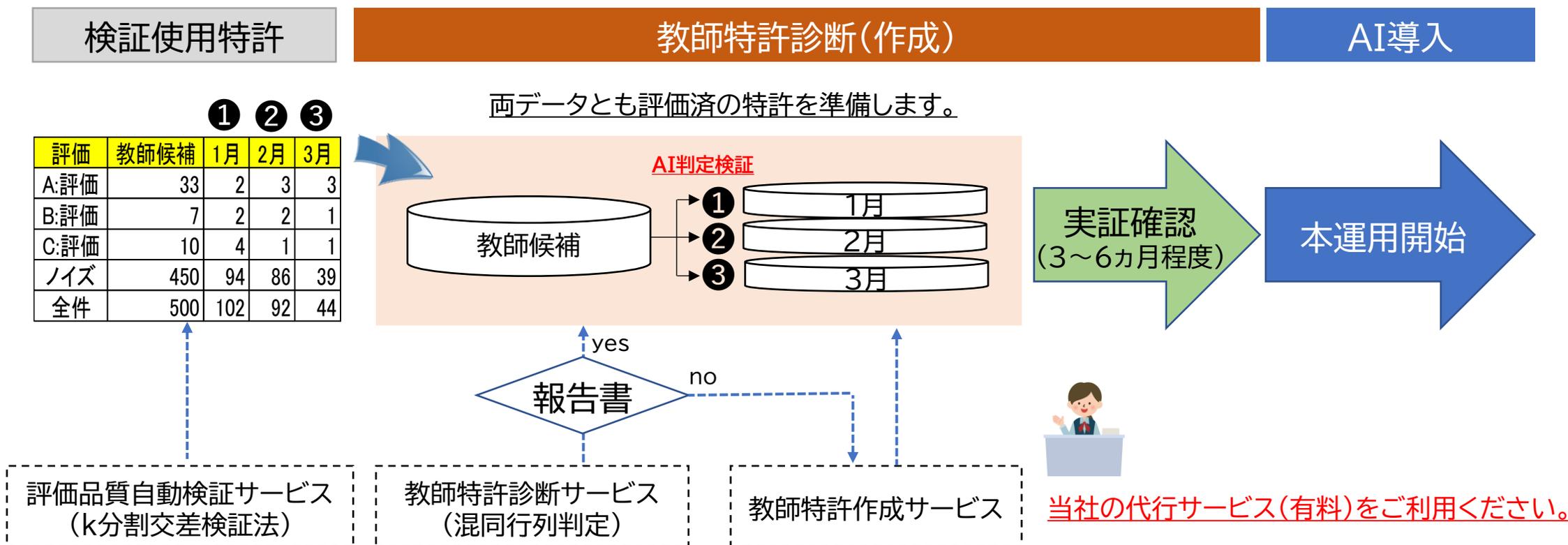


(AIが自らの基準で何かしら人知れぬ処理をブラックボックスで行っている訳ではありません)

# 初めてAIを活用する(推奨方法)

SDI調査テーマでは、過去からの評価付け特許が存在していますので、AIをまず使ってみることが比較的容易にできます。具体的には、次の方法がお勧めです。

- ①. 判定特許は、既に評価が終わっている直近の3～6か月分を対象にします。
- ②. 教師特許(評価付、ノイズ含)は、それ以前の6か月～1年分を「教師候補」として準備します。



# SDI検証を混同行列で確認

SDI調査の社内評価データと月別AI検証データを使って混同行列を作成します。  
目的は、偽陰性判定の発生状況を確認します。偽陰性判定された特許は評価の確認を行います。

公開SDI '21/1月		AI判定	
		サーチ	ノイズ
社内評価	抽出	0(0%)	1(100%)
	ゴミ	84(30%)	196(70%)

公開SDI '21/4月		AI判定	
		サーチ	ノイズ
社内評価	抽出	13(100%)	0(0%)
	ゴミ	118(28%)	282(72%)

公開SDI '21/2月		AI判定	
		サーチ	ノイズ
社内評価	抽出	9(100%)	0(0%)
	ゴミ	122(30%)	285(70%)

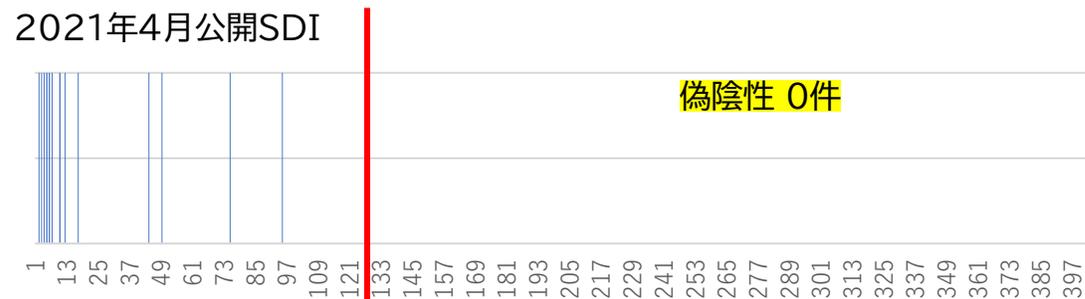
公開SDI '21/5月		AI判定	
		サーチ	ノイズ
社内評価	抽出	7(100%)	0(0%)
	ゴミ	83(28%)	210(72%)

公開SDI '21/3月		AI判定	
		サーチ	ノイズ
社内評価	抽出	5(100%)	0(0%)
	ゴミ	114(30%)	266(70%)

公開SDI '21/6月		AI判定	
		サーチ	ノイズ
社内評価	抽出	4(100%)	0(0%)
	ゴミ	82(29%)	199(71%)

# 関連特許の判定順位を確認

SDI調査におけるAI判定の検証結果を可視化する1つの方法で、“青線”が関連特許の判定順位です。偽陰性特許は評価内容を確認します。偽陰性がなければこの教師特許で本番移行することができます。



# 混同行列を使った判定結果評価法

AI機能を使って判定した場合の検証方法について説明します。混同行列を用いて確認する方法です。縦軸を社内評価値、横軸をAI判定結果としたマトリクスを作成します。

両判定が一致したところが正解です。食い違ったところが「偽陽性」「偽陰性」となり教師特許の問題によるものか、評価上問題がある特許か、AI処理の限界超えによるものかの確認を行います。

当社では、クリアランス調査の場合、関連特許の判定は**100%**を目指します。

		AI判定			
		サーチ		ノイズ	
社内評価	関連	10	100%	0	100%
	参考	18	90%	2	10%
	ノイズ	57	19%	243	81%

社内評価とAI判定のどちらが正解なのか**再評価の対象**となります。社内評価が正解の場合は、教師特許の見直しが必要です。

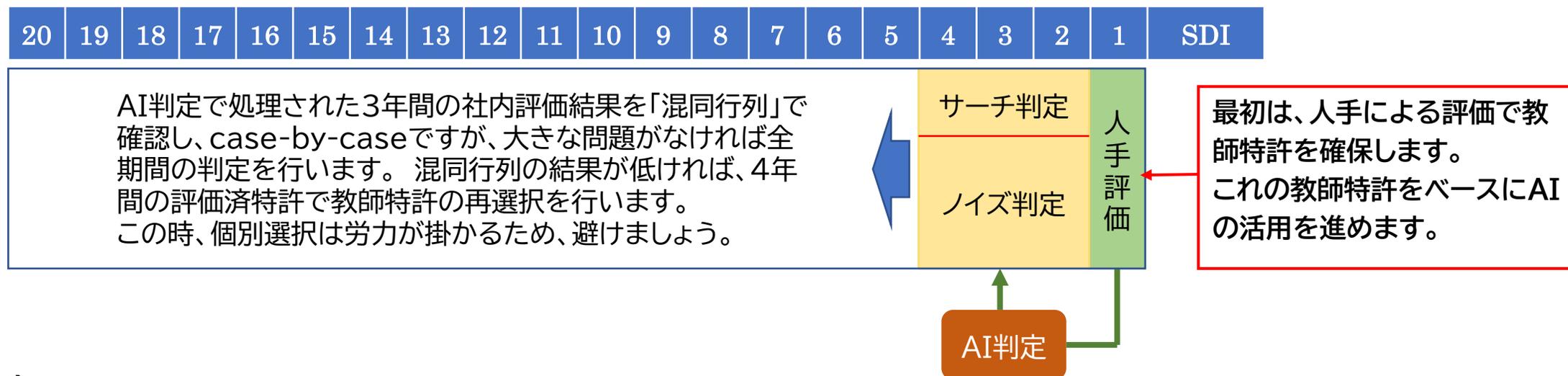
教師特許に類似し判定順位が高いにもかかわらず、社内でノイズとした特許です。見直し優先度の高いノイズ特許の再評価で、**評価品質を効率的に確認**できます。

境界値:35%

AIを活用した特許調査の**効率化を実現する対象**となります。

# 過去遡及調査のAI活用

過去20年間の遡及調査(クリアランス, 動向調査etc.)では、基本的に教師特許がありません。この場合、一定期間(例えば、直近1年間分等)は人手(※)による評価が必要です。評価済みの特許をサーチ教師/ノイズ教師として活用します。(但し、他の類似の特許調査結果がある場合は、それを教師とすることができます。)



(※)  
THE調査力AIには、キーワードを活用したシン・ハイライトソート機能があります。この機能を活用して人手による評価をより効率的に行いことができます。

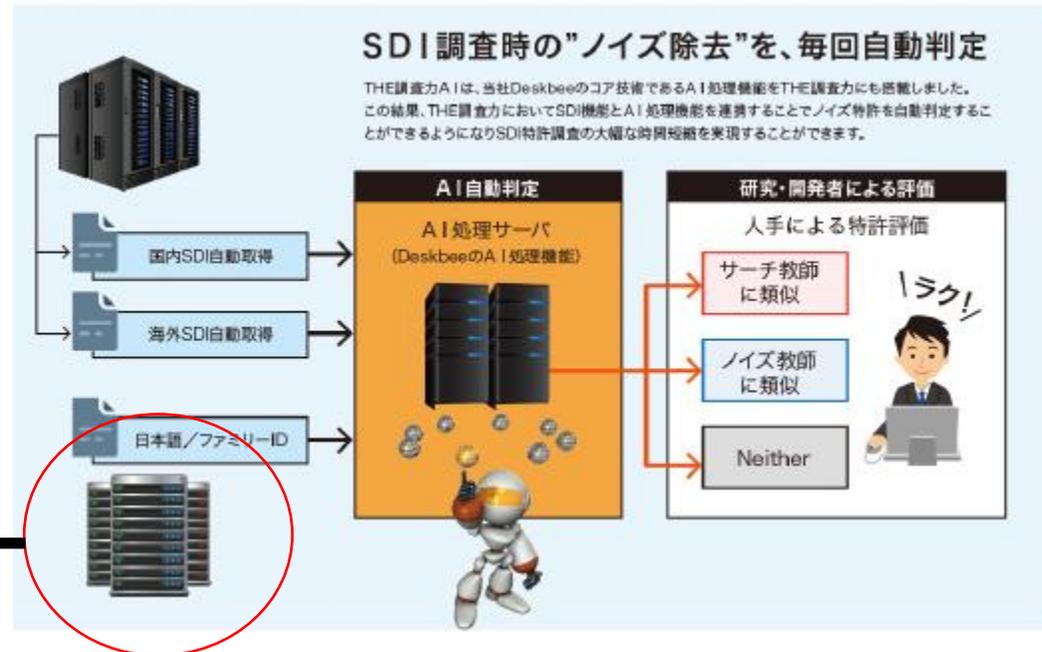
# 海外特許調査のAI活用も可能

海外特許調査には、全世界特許和文データベース「Japio-GPG/FX」の日本語活用がお勧めです。このデータベースは、既に和文が蓄積されたデータベースなので短時間でデータ取得することができること、精度が高く全世界の特許を同一の翻訳エンジンで翻訳しているため訳質が均質化されていることが特長です。海外特許調査のAI判定処理にはこの和文データをご活用いただけます。

THE調査力に高精度の日本語・ファミリー公報がまとめて取り込める



Japio-GPG/FXの日本語でAI判定、そのまま評価も日本語で



# 評価品質自動検証機能

特許調査は、件数が多く負担がかかるため、一度の評価で調査を終えます。また分担による評価のブレやSDI調査での評価軸の変更など、再確認できるならば見直しの必要性もあります。

そこで当社では、“k分割交差検証法”と“混同行列”の作成を完全自動化した「評価品質自動検証」を業界で唯一AIで実現しました。

対象特許を均等に4分割する第一工程

1群教師	2群	3群	4群
1群	2群教師	3群	4群
1群	2群	3群教師	4群
1群	2群	3群	4群教師

マクロ付診断結果データ出力の第二工程



診断結果を利用者が再確認する第三工程

		AI判定			
		サーチ		ノイズ	
ユーザ 評価	A	9	90%	1	10%
	B	18	90%	2	10%
	ゴミ	57	19%	243	81%

※次ページで実際の実出力例をご参照いただけます。

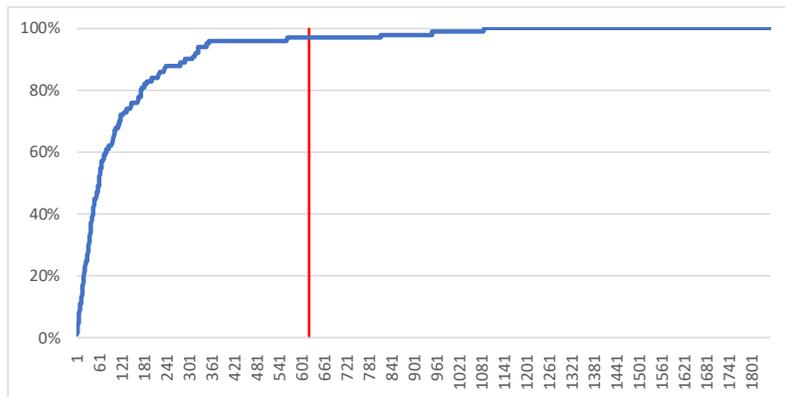
評価ランクを見直し  
関連特許がないか？

# 評価品質自動検証の出力例

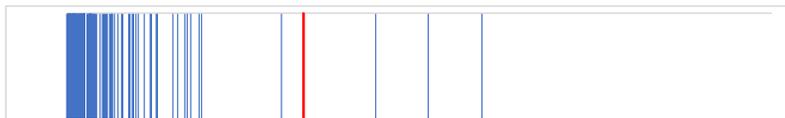
## ■ 混同行列

ユーザ評価	AI判定	AI判定	
		サーチ	ノイズ
A	61件	100%	0件 0%
B	12件	100%	0件 0%
C	17件	85%	3件 15%
ゴミ	531件	30%	1,229件 70%

## ■ 再現率飽和グラフ



## ■ サーチ案件出現位置グラフ



## ■ サーチ・ノイズ閾値

	件数
サーチ	93件
ノイズ	1,760件
S/N閾値	621件

## ■ 偽陰性案件リスト

人間がサーチ側（重要）と評価したにも関わらず、AIはノイズに近い（非重要）と判定した案件を、ノイズに近い側から順に列記しました。本当に重要な案件なのかどうか再評価してみてください。もしこれらが本当に重要な案件であったとしても、現状の第3世代AIにこの案件群を重要と認識させることは難しいとされます。ハイライトソートの併用によりSDI案件の抽出漏れを防止する方法をお勧めします。

右のボタンをクリックすると、THE調査力の抄録画面を使って、これらの案件を  
[THE調査力抄録表示](#)  
 査読することができます。

順位	出願番号	権利者・出願人名	発明の名称	国コード	公報種別コード
1089	特願2017-000079	アサヒビール株式会社	ビール様発泡性飲料及びそ	JP	B9
950	特願2002-219787	株式会社帯広ビール	発泡飲料	JP	B9
812	特願2020-215050	サントリーホールディングス株	発酵ビールテイスト飲料	JP	A

便利

## ■ 偽陽性案件リスト

人間がノイズ側（非重要）と評価したにも関わらず、AIはサーチに近い（重要）と判定した案件を、サーチに近い側から順に列記しました。判定に間違いがないかどうか、数十件程度を再評価してみてください。

右のボタンをクリックすると、THE調査力の抄録画面を使って、最大100件まで  
[THE調査力抄録表示](#)  
 これらの案件を査読することができます。

順位	出願番号	権利者・出願人名	発明の名称	国コード	公報種別コード
11	特願2014-031041	アサヒビール株式会社	発酵麦芽飲料	JP	B9
22	特願2011-286375	麒麟麦酒株式会社	ホップ香気を強調した発酵	JP	B9
25	特願2013-271064	アサヒビール株式会社	発酵麦芽飲料	JP	B9
27	特願2013-271065	アサヒビール株式会社	発酵麦芽飲料	JP	B9
33	特願2013-271063	アサヒビール株式会社	発酵麦芽飲料	JP	B9
41	特願2020-068493	サントリーホールディングス株	ビールテイスト飲料およびそ	JP	A
49	特願2014-031049	アサヒビール株式会社	発酵麦芽飲料	JP	B9

便利

上位から数十件程度を確認

# Q&A

Q. 教師特許は、何件くらいあればいいですか？

A. 100件程度～1,000件(無効資料調査の場合は、サーチ教師1件～ノイズ教師は30件以上)もあれば十分です。サーチ・ノイズ両教師は、実際の調査での比率が望ましいですね。SDI調査では、最近の一定期間の調査結果をそのまま教師にできます。

Q. 調査データにいろいろな種類(分野)のデータが入っていますが、問題ないですか？

A. 基本的に類似分野のデータ単位でAI処理する方が判定精度が高くなります。例えば、(材料と製造方法と応用品)が混在しているケースなどでは、材料の教師が多い場合に他はノイズ判定されることとなります。

Q. 偽陰性がでた場合、どのようにすればいいですか？

A1. まず偽陰性特許の評価を確認します。変更がなければ教師特許の見直しが必要です。但し、1件ごとの見直しは避けましょう。評価ランクの選択を変更する、期間を変えるなどでサーチ教師を見直します。

A2. 評価ランクが“参考”部類の特許は、偽陰性率が高くなる傾向にあります。”参考”評価をつけた選択基準が担当者によってまちまちだから教師とした特許もいろいろな視点が存在しているケースです。

Q. 海外特許での活用において注意点はありますか？

A1. JapioGPG/FXの翻訳では、大別(EJ)と(CJ)のエンジンがあります。英語文献と中国文献はそれぞれの範囲内の特許を教師として分けてAI判定を掛けることがお勧めです。

A2. 英文特許判定で偽陰性がでた場合、それが少件数のマイナー国の特許である時は最初からAI判定から除外して人手評価をお勧めします。

Q. 重要キーワードがテキスト中にあるにもかかわらず、判定順位が低くなるのは何故ですか？

A. AIには利用者が何が重要キーワードなのかを判別するデータがありません。またAI処理で継続した100%はあり得ません。このため、当社独自の機能として“シン・ハイライトソート”を標準搭載しています。重要キーワードを使いたい場合はAIクロス集計が有効です。

THE調査力AIの操作方法は、AI機能のトライアル操作マニュアルをご覧ください。