

ソフトバンク国産AI Sarashinaの最新状況と知財業務への活用可能性評価

エグゼクティブサマリ

本報告の結論を先に述べると、**Sarashinaは「知財業務の全面自動化」には向きませんが、厳格な人手審査を前提にした「知財業務支援基盤」としては有力候補**です。特に、要約・翻訳、文書構造化、社内ナレッジ検索、出願ドラフトのたたき台作成、契約条項の抽出・比較、先行技術調査のクエリ展開や一次整理といった用途では、Sarashina API、Sarashina2.2-Instruct、Sarashina2.2-Vision-3B、Sarashina2.2-OCRの組み合わせが実務的です。一方で、**侵害リスク評価、最終的な先行技術判断、登録可否判断、クレーム最終化、法的見解の提示**を単独で任せるのは、公開資料の範囲では不適切です。SB Intuitions 自身も複数モデルについて「限定的な安全学習」「不正確・偏った出力の可能性」を明示しており、APIの公開資料でも料金・SLA・ログ保持・削除期限などの契約実務上重要な条件は未公開です。 ¹

商用・法務利用の可否は、モデル系統ごとに大きく異なります。公開OSSのうち、Sarashina1/2、Sarashina2.2、Sarashina2.2-Instruct、Sarashina2-Vision、Sarashina2.2-Vision-3B、Sarashina2.2-OCRにはMITライセンスが確認できます。他方、Sarashina2-8x70B、Sarashina2.1-1B、Sarashina-Embedding-v1-1B、Sarashina-Embedding-v2-1Bには**Sarashina Model NonCommercial License Agreement**が適用され、商用利用に制限があり、さらに禁止行為ポリシーでは**“legal”を含む legally required work への Output Data 利用が禁止**されています。したがって、**非商用ライセンス系のOSSを、そのまま知財・法務の本番業務に使うのは不可**というのが、本報告で最も重要な実務上の注意点です。 ²

最新の製品化状況は、研究モデル段階から企業提供段階へ進んだと評価できます。2024年には7B/13B/65B/70Bの事前学習モデル公開、2024年11月には8x70BのMoE公開、2025年には2.1/2.2系の小型高性能モデル、Vision系、Embedding系が拡充され、2025年11月に法人向け**Sarashina API**が開始されました。さらに2026年4月には日本国内データセンターでソフトバンクが管理・運用する**Cloud PF Type A**が開始され、2026年6月から順次**Sarashinaを活用した生成AIサービス**を提供すると発表されています。機密性・データ主権を重視する知財部門にとって、この「API商用提供」と「国内ソブリンクラウド提供」の二本立ては強い材料です。 ³

ただし、**公開情報の粒度は研究寄り**で、**エンタープライズ契約に必要な粒度は不足**しています。公開資料で確認できるのは、入力・出力データをモデル学習に利用しないこと、カスタマイズは現時点で未対応であること、セキュリティ詳細や各モデルの料金は問い合わせ対応であること、Cloud PF Type Aが国内データセンター運用であること、という水準です。**ログ保持期間、削除証跡、監査対応、再委託先、SLA、障害時補償、モデル変更通知、閉域接続、VPC/専用線、APIコンテキスト長・レート制限・トークン単価**は、公開ページ上では未確認です。知財用途で本格導入する場合、PoC前またはPoC契約時に必ず確認すべきです。 ⁴

結論

現時点の導入判断は、「**限定導入は推奨、本番全面依存は非推奨**」です。より実務的に言えば、**知財部門向けの第一選択肢はSarashina APIまたはCloud PF Type A上のSarashina利用であり、OSSを使う場合はMITライセンス系に限定し、非商用ライセンス系は本番知財業務から外す**、という方針が妥当です。これは、Sarashina APIが法人向けの商用サービスとして提供され、入力・出力を学習に使わないと明示している一方、Embedding系OSSや8x70Bには商用制限とlegal work制限があるためです。 ⁵

知財業務に即して採否を分けると、**要約・翻訳・契約レビュー補助・文書構造化・RAG検索補助は採用候補、先行技術調査の一次整理や出願ドラフト補助は条件付き採用、侵害リスク評価・登録可否判断・最終法律意見・最終クレーム確定は人間主導に限定**が妥当です。特許・商標の公式探索基盤としては J-PlatPat が公式に提供されており、AI はその前段・周辺処理に使うべきです。また、文化庁の「AIと著作権に関する考え方」は、AIと著作権の関係についての考え方を示す一方で、**法的拘束力を持たず、個別事案での確定評価ではない**と明示しています。つまり、著作権・学習データ・生成物類似性の論点は、公開資料だけで完全には解消されません。 6

本報告としての最終的な推奨は、「**知財コパイロット**」としての導入を推奨し、「**知財オートパイロット**」としての導入は見送る」です。具体的には、社内文書・公開公報・契約台帳・過去 OA 応答・社内審査基準を RAG 化し、Sarashina mini または Sarashina2.2-Instruct 系で草案生成・要約・比較を行い、特許庁・文化庁・J-PlatPat 等の一次ソースへ常に戻れる設計にすべきです。Vision/OCR 系は、図表・縦書き・複雑レイアウトの処理が必要な知財文書で有効です。 7

最新状況と技術・提供形態

Sarashina の公開情報を時系列で整理すると、研究開発基盤から実務サービスへと段階的に進化してきたことが分かります。2024年6月に 7B/13B/65B の Sarashina1/2 系が公開され、その後 2024年8月に 70B、2024年11月に 8x70B の MoE、2025年に 2.1-1B、2.2-0.5B/1B/3B、Instruct、Vision、Embedding、2026年3月に OCR、2025年11月に Sarashina API、2026年6月から Cloud PF Type A 上の生成AIサービス提供へと進みました。ソフトバンク IR では、**2024年度に 4,600億パラメーターの基盤モデル Sarashina が完成し**、その知見を活用した **Sarashina mini** の商用化を進めていること、さらに **次世代の 1兆パラメーター規模 LLM** 開発を目指していることが説明されています。 8



上図の流れが示す通り、最新の「**企業導入可能なSarashina**」は、研究用オープンウェイトそのものではなく、**Sarashina mini** を中核とする **Sarashina API** と、**国内データセンター運用の Cloud PF Type A 上で順次提供される生成AIサービス**です。Sarashina API は Chat Completion API と Embeddings API で構成され、2025年11月28日に開始されました。ソフトバンクと SB Intuitions は 2025年6月から約2万人の社員トライアルを行って改良してきたと説明しています。2026年4月発表では、Cloud PF Type A において Sarashina を使うことで、**機密情報や顧客データを連携した高精度生成AIを、データ主権を備えた環境で利用できる**とされています。 9

公開済みの主要モデル・サービスを、知財実務の観点で整理すると次の通りです。

系列	公開・提供状況	主な仕様	ライセンス/商用条件	知財実務上の含意
Sarashina1/2 7B・13B・65B/70B	Hugging Face 公開済み	Sarashina1 は GPT-NeoX、Sarashina2 は Llama 構造。Sarashina2 は日・英・コード、2.1T トークン、語彙 102,400。70B は事後学習なしで日本語QA 平均 83.36。 ¹⁰	MIT。商用可。70B は対話用ではない。 ¹¹	オフライン検証や日本語特化の基盤モデル評価に有用。ただし Instruct ではないため、そのまま知財対話には不向き。
Sarashina2-8x70B	Hugging Face 公開済み	Sarashina2-70B を sparse upcycling で MoE 化。465B params。BF16 推論に 16x H100 または 16x A100 80GB を要求。 ¹²	NonCommercial。商用制限あり。 ¹³	研究用途では魅力があるが、法務・知財本番にはそのまま使えない。インフラ要件も重い。
Sarashina2.1-1B	Hugging Face 公開済み	10T の日英 Web コーパス + 1T の日本語重視学習。1B params。 ¹⁴	NonCommercial。商用制限あり。 ¹⁵	小型だが本番知財業務には不適。
Sarashina2.2 0.5B/1B/3B と Instruct	Hugging Face 公開済み	10T 日英コード + 合成データで数学/コード強化 + 少量の応用データ。公開コレクションは 0.5B/1B/3B。Instruct は ELYZA-tasks-100 と MT Bench を公開比較。 ¹⁶	MIT。商用可。Instruct は安全学習が限定的。 ¹⁶	OSS で本番候補になり得る中核系列。 特に 3B Instruct は PoC の第一候補。
Sarashina2-Vision 8B/14B、Sarashina2.2-Vision-3B	Hugging Face 公開済み	14B は Sarashina2-13B + Qwen2-VL-7B image encoder。3B は Sarashina2.2-3B-Instruct + SigLIP。日本語図表/文書画像で高性能。 ¹⁷	MIT。商用可。 ¹⁸	商標図形、特許図面、契約 PDF、証拠画像の理解補助に有効。
Sarashina2.2-OCR	Hugging Face 公開済み	日本語/英語文書向け end-to-end OCR。縦書き・表・数式・Markdown/HTML/LaTeX 再構成。説明部では 3B-parameter とされる一方、HF メタデータは 4B params。 ¹⁹	MIT。商用可。 ²⁰	複雑 PDF の知財ワークフロー前処理に極めて有用。公開情報にパラメータ数の不整合があるため、正式仕様は確認が必要。

系列	公開・提供状況	主な仕様	ライセンス/商用条件	知財実務上の含意
Sarashina-Embedding v1/v2	Hugging Face 公開済み	v1/v2 とも 8,192 tokens、1792 次元。JMTEB で高スコア、v2 は 2025年7月時点で平均 76.38。 ²¹	NonCommercial。商用利用には別途連絡が必要。 ²¹	性能は魅力的だが、そのまま本番知財 RAGには使えない。商用許諾が前提。
Sarashina API / Cloud PF Type A	法人向け 商用提供	Sarashina mini と接続する API。Chat Completions と Embeddings。2026年6月からは国内データセンター運用の Cloud PF Type A でも順次提供。 ⁹	商用サービス。料金・詳細モデル・セキュリティ詳細は要問い合わせ。 ²²	知財部門の第一候補。契約条件の詰めが必要。

アーキテクチャと学習データの公開粒度は、モデル世代によってかなり差があります。Sarashina1/2 については、Sarashina1 が GPT-NeoX、Sarashina2 が Llama 構造、Sarashina2 は日本語・英語・コードの 5:4:1 混合、英語に SlimPajama、コードに starcoderdata、日本語は CommonCrawl 由来の独自構築データを CCNet・HojiChar・MinHashLSH で整備したことまで公開されています。さらに、**SlimPajama からは著作権侵害のおそれがあるとして書籍コーパスを除去したと明記されています**。対して、4,600億パラメーター基盤モデル Sarashina 本体の詳細アーキテクチャ、学習トークン数、コンテキスト長、推論コスト、API で実際に露出する Sarashina mini の詳細仕様は、公開資料では十分に確認できません。IR では **Sarashina mini は蒸留した 700億パラメーター**と説明されますが、製品ページではその数値が示されていません。²³

推論速度・コストの公開値は限定的です。API単価やトークン単価、SLA ベースの応答時間は未確認です。公開されている定量情報としては、Sarashina2 の日本語トークナイザ効率が高く、例示では Llama3/Qwen2 が同一日本語文で Sarashina2 の約 2.5 倍のトークン数となり、まとまった日本語テキストでも他モデル比で約 1.5~1.7 倍のトークン数差があるとされています。これは日本語処理における計算効率上の利点を示唆しますが、**APIレイテンシや実運用コストの代替にはなりません**。8x70B だけは Required Hardware が公開されており、BF16 推論に 16x H100 または 16x A100 80GB を要します。²⁴

ライセンス・セキュリティ・契約上の論点

知財実務で最初に確認すべきなのは、**Sarashina は「一枚岩の利用条件」ではない**という点です。MIT ライセンスのモデル群と、NonCommercial ライセンスのモデル群が混在しています。MIT 群には、少なくとも Sarashina1/2、Sarashina2.2、Sarashina2.2-Instruct、Sarashina2-Vision、Sarashina2.2-Vision-3B、Sarashina2.2-OCR が含まれます。他方、Sarashina2-8x70B、Sarashina2.1-1B、Sarashina-Embedding-v1-1B、Sarashina-Embedding-v2-1B は NonCommercial です。**知財部門が社内 PoC から本番展開へ進める際、この境界を曖昧にすると、技術評価は通っても契約・利用条件で止まる可能性があります**。²⁵

とりわけ重要なのは、NonCommercial 群の禁止行為ポリシーです。同ポリシーでは、**他モデル改善への利用制限に加え、“legal”を含む legally required work への Output Data 利用禁止**が明示されています。知財実務のうち、特許・商標・著作権の調査報告、侵害リスク評価、契約レビュー結果、出願補正理由の整理などは、通常「法務・知財の正式業務」に近接します。したがって、**NonCommercial 群の公開モデルを使って本番知財業務に成果物を出すのは、公開ライセンス上かなり危険**です。もし Sarashina-Embedding-v2-1B のような性能上有望なモデルを使いたいなら、**SB Intuitions との個別商用許諾が前提**と解釈すべきです。²⁶

一方で、Sarashina APIは商用サービスとして提供されており、公開FAQでは顧客の入出力データをモデル学習に利用しないこと、カスタマイズは現時点では未対応であること、セキュリティ・データ保護対策の詳細は問い合わせであること、各提供モデルの詳細や料金は問い合わせであることが示されています。これは、エンタープライズ向けには前向きな材料ですが、逆にいえば知財用途で必須になる保存期間、削除条件、ログマスキング、監査ログ、障害時復旧、モデル更新通知、下請け・再委託先、閉域接続、データ持ち出し禁止などは公開されていません。したがって、導入可否は「技術評価」だけでなく「契約評価」で決まります。 27

セキュリティ面では、SB Intuitionsは情報セキュリティポリシーでISMS適用範囲のシステム運用業務について年1回以上のリスクアセスメント、法令・契約順守、内部監査と継続的改善を掲げています。また個人情報保護方針では、組織的・人的・物理的・技術的安全管理措置を掲げています。他方、同方針は会社全体として海外で個人データを取り扱う場合があることにも触れており、Sarashina API/Cloud PF Type Aに限った再委託先や越境移転の有無は未確認です。Cloud PF Type A自体は日本国内データセンターにOracle Alloyを導入し、データの所在・管理・運用を日本国内で完結させると説明されているため、高機密な知財案件ではCloud PF Type Aの方が説明しやすい設計です。 28

学習データの出所と権利関係については、公開情報から読み取れる範囲と、未確認の範囲を分けて考える必要があります。読み取れる範囲としては、Sarashina2では日本語にCommonCrawl由来の独自コーパス、英語にSlimPajama、コードにstarcoderdataを用い、書籍コーパスは除外したことが確認できます。Sarashina2.2では日英コードのWebコーパス、数学・コーディング用の合成データ、応用タスク向上のための少量データが使われたことが示されています。未確認の範囲としては、日本語コーパスの具体的サイト一覧、著作権者対応手順、サイト単位のオプトアウト運用、2.2の合成データ生成元、アプリケーションタスク用データの権利処理です。文化庁の「AIと著作権に関する考え方」は、著作権法第30条の4等の基本的考え方を整理していますが、同文書自体は法的拘束力を持たず、個別生成AIについての確定的評価ではないと明言しています。したがって、「国産だから権利問題が解決済み」とまでは言えません。 29

もう一つ注意すべき点として、SB Intuitionsのクローラーポリシーでは、利用者がSarashinaに質問するとSearchBotがウェブサイトを参照して推論に使う場合があること、ただし参照した情報をAI開発には使用しないことが示されています。これがSarashina APIの標準機能として常時有効なのか、企業向け環境で無効化・制御できるのかは、公開資料だけでは未確認です。知財調査では、参照先の真正性、検索範囲、記録保存、引用元明示が重要になるため、Webブラウズ機能の有無と制御可否は必須確認項目です。 30

性能・限界・知財業務シナリオ別評価

Sarashinaの性能は、日本語QA、翻訳、日本語文書理解、埋め込み検索で特に強みがあります。Sarashina2-70Bは、日本語QAの6指標平均で83.36を示し、同記事内の比較ではSwallow-70BやLlama系を上回っています。英日・日英翻訳でも平均25.48と比較上位です。また、Sarashina2は日本語トークナイザ効率が高く、日本語特有の長い系列を短く扱えるため、日本語の知財文書処理では理論上の計算効率面の利点があります。 11

小型系では、Sarashina2.2が「小さくても強い」点が重要です。Sarashina2.2-1Bのカードでは、3段階学習により数学・コードを強化し、3Bモデルは日本語QAの一部で旧7B/70Bを上回ると説明されています。Instruct系では、1BモデルでもELYZA-tasks-100で2.88、3Bでは3.75、Japanese MT Benchでも3Bが6.51を示しており、同サイズ帯で実用的な対話性能を持つことが分かります。特許明細書ドラフト補助、要約、条項比較、拒絶理由の平易化説明といった「人が確認する前提の知財対話」では、2.2-3B Instructがコストと性能のバランスに優れた候補です。 16

文書画像系では、Sarashina2.2-Vision-3BとSarashina2.2-OCRが知財用途に刺さります。Vision-3BはBusinessSlide VQA、Heron-Bench、JDocQA、JMMMUなど日本語系ベンチマークで小型帯として高いスコアを示しており、OCRは縦書き日本語、複雑スライド、表、数式をMarkdown/HTML/LaTeXに再構成できま

す。OCRのベンチマーク結果では、VJRODaでSarashina2.2-OCRがCER 22.6、BLEU 79.9を示しており、日本語の複雑レイアウト読取に強みがあります。これは、紙PDFの契約書、審決・判決PDF、古い公報PDF、図表を含む明細書、証拠資料の構造化に直接役立ちます。³¹

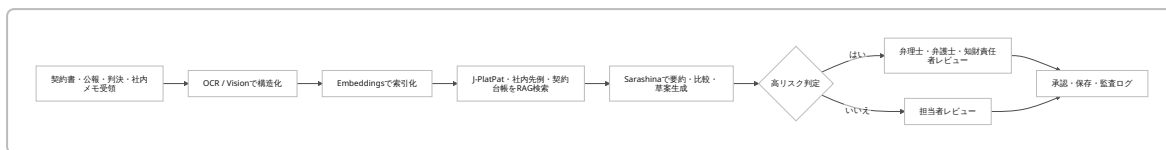
RAGには、Embeddings APIか、別途商用許諾を得た埋め込みモデルが必要です。公開OSSの埋め込みモデルはv1/v2とも8,192トークン、1792次元で、JMTEBでは高スコアです。v1は平均75.50、v2は2025年7月評価で平均76.38を示し、OpenAI text-embedding-3-largeを上回る公開比較もあります。外部のNAACL 2025 Industry論文でも、Sarashina系埋め込みはSOTA比較対象として扱われています。ただし前述の通り、**OSSのEmbedding v1/v2はNonCommercial**であるため、知財本番では**Sarashina APIのEmbeddings API**を使うか、別途商用条件を確定させる必要があります。³²

安全性・誤情報・バイアスについては、過信は禁物です。Sarashina2.2-Instructのモデルカードは、**安全学習が限定的であり、意味のない列、不正確な記述、偏向的/不快な出力があり得る**と明示しています。事前学習モデル群では、そもそもinstruction tuning未実施のため、対話的タスクにはそのまま向かないとされています。一方で、SB Intuitionsは日本語安全性境界テストを公開し、2.2-3B instructが同程度サイズのモデルより高いスコアを示したと報告し、さらにAPIでは**Sarashina2 guard**を使えると案内しています。つまり、**安全性努力は進んでいるが、法務・知財のhigh-stakes判断を任せられるほど公開保証されているわけではない**というのが実態です。³³

知財業務シナリオごとの評価をまとめると、以下の通りです。

業務シナリオ	推奨構成	適合度	主なリスク	判断
特許・商標・著作権調査の一次整理	Sarashina API + Embeddings API + J-PlatPat/社内DB RAG	高い	公式DBを代替できない。検索漏れ、引用元欠落、法的評価の飛躍。 ³⁴	条件付き可
先行技術調査	J-PlatPat 検索を主、AIはクエリ展開・要約・クラスタリング補助	中程度	exhaustive recallを保証できない。分類付与やclaim chartは人手必須。 ³⁵	条件付き可
出願書類作成	Sarashina2.2-InstructまたはSarashina APIで草案生成、人間がクレーム最終化	高い	クレーム幅の過不足、サポート要件逸脱、虚偽・過剰効果表現。モデルの不正確出力。 ³⁶	補助用途で可
侵害リスク評価	条文・判例・公報の要点抽出まで。最終意見は弁護士/弁理士	低い	事実認定・クレーム解釈・均等論・商標類否で誤誘導リスクが高い。 ³⁷	単独利用は不可
契約レビュー	条項抽出、差分比較、要注意条項の一次フラグ付け	高い	誤読・条項間整合性欠落・業法/裁判実務の見落とし。公開安全保障不足。 ³⁸	補助用途で可
要約・翻訳	Sarashina API、2.2-Instruct、70B系	非常に高い	固有名詞誤訳、権利範囲を変える要約、出典不一致。 ³⁹	強く推奨
PDF/画像の構造化	2.2-OCR + Vision-3B	非常に高い	OCR誤読、表崩れ、図面番号取り違い。 ⁴⁰	強く推奨

知財ワークフローへの統合は、次のような「人間審査を組み込んだ構造」が適切です。公開されている Sarashina API、Embeddings API、Vision/OCR の能力と、法務・知財判断を自動委任すべきでないという制約を前提にした推奨構成です。 41



推奨導入方針とPoC設計

推奨導入方針は、「API/サブリンククラウドを本命、MIT OSS を補完、NonCommercial OSS は本番対象外」です。知財部門にとって重要なのは、精度だけでなく、情報の出所管理・削除統制・アクセス統制・監査可能性です。そのため、第一段階の PoC は **Sarashina API** を主系にし、国内データ主権要件が厳しい案件や将来本番では **Cloud PF Type A** を候補に置くのが妥当です。どうしても文書前処理を自社内で完結したい場合のみ、MIT の **Sarashina2.2-OCR** と **Sarashina2.2-Vision-3B** を限定的に使う二層構成が現実的です。 42

この方針に沿うなら、PoC で検証すべきなのは「一般ベンチマークの点数」ではなく、**知財部門のゴールデンセットでの再現性**です。公開ベンチマークは有用ですが、知財実務では、調査漏れ 1 件、条項誤読 1 件、クレーム表現の 1 文の曖昧化が大きな影響を持ちます。したがって、PoC の評価指標は、**正答率**だけでなく、**引用根拠付与率、検索再現率、幻覚率、レビュー工数削減率、機密情報取り扱い違反件数**を含めるべきです。 43

PoC設計案

要件	評価指標	期間	担当
要約・翻訳補助の有効性検証	人手採点 5段階平均、事実誤り率、引用付き要約率、レビュー時間削減率	2週間	知財部門、翻訳担当、情報システム
先行技術調査の一次整理	ゴールデンセット公報に対する Recall@20、Hit@5、分類提案の妥当性、根拠URL/文献提示率	3週間	特許担当、検索専門員、弁理士
契約レビュー補助	重要条項抽出 F1、差分検出率、要注意条項フラグの Precision、幻覚条項率	2週間	法務、知財契約担当
OCR/Vision 前処理	CER、表再現率、図番号誤読率、Markdown/HTML 変換再利用率	2週間	情報システム、DX推進、知財事務
セキュリティ・運用性	PII/営業秘密の誤送信 0件、アクセス権逸脱 0件、ログ設計の妥当性、削除手順検証	並行 2週間	情シス、CISO/セキュリティ、法務
全体採否判定	工数 20~40%削減、重大誤りゼロ、引用根拠付与必須工程で 95%以上、利用規約適合	1週間	知財責任者、法務責任者、情報システム責任者

PoC の前提データは、**社外秘をいきなり投入しない**ことが重要です。第一段階では、公開公報、公開審決、ひな形契約、既に社外開示済みの技術資料、匿名化した過去案件を使い、第二段階でのみ機微案件に拡張するのが安全です。Embeddings API を使う場合も、**案件番号・発明者名・相手方名・未公開発明の表題**などの機微識別子は、初期 PoC ではマスキングした上で評価すべきです。公開 FAQ では入力・出力が学習に使われないことは確認できますが、保持期間や監査証跡は未公開だからです。 44

契約条項としては、次の最低限を推奨します。これは、公開資料で未確認の点を契約で埋めるための実務提案です。公開ページで「問い合わせ」とされている事項は、契約本文または別紙運用仕様に落とし込むべきです。 ⁴⁵

条項案	最低要求水準
顧客データの二次利用禁止	入出力、埋め込み、ログ、メタデータを学習・再学習・評価・第三者提供に使わないことを明示
データ所在と再委託	国内保管、再委託先一覧、越境移転の有無、変更時通知義務
保持と削除	本文・ログ・キャッシュ・バックアップの保持期間、削除期限、削除証跡の提供
セキュリティ	暗号化、アクセス制御、監査ログ、脆弱性対応、インシデント通知期限
モデル変更管理	基盤モデル更新時の事前通知、性能影響説明、ロールバック手段
SLA/SLO	可用性、応答時間目標、障害通知、補償、メンテナンス窓口
知財条項	出力物の権利帰属、第三者権利侵害申立て時の協議手順、学習データ由来紛争時の責任分担
監査・説明責任	監査質問権、第三者認証資料、セキュリティ説明書の提供
閉域接続・分離環境	必要に応じて閉域/VPN/専用線/VPC 相当の選択肢を明示
終了時対応	ベクトルDB・ログ・キー・監査情報の返還/削除、エクスポート性、代替移行支援

未解決事項とソフトバンクへの確認項目

公開情報には強みが多い一方、知財本番導入では未解決事項がかなり残っています。以下は、PoC 前に必ず確認すべき項目です。**未確認**とした事項は、公開ページでは確認できなかったものです。 ⁴

確認項目	公開情報の状況
Sarashina API で実際に使われるモデル名・バージョン・パラメータ数	未確認 。API発表は「Sarashina mini」とのみ明示し、IR では700億と説明。 ⁴⁶
API のコンテキスト長、出力トークン上限、レート制限	未確認
API 利用料、初期費用、最低利用料金、従量課金方式	未確認 。料金は問い合わせと明記。 ⁴⁷
SLA/SLO、可用性保証、補償条件	未確認
ログ保持期間、ログ項目、削除条件、バックアップ保持	未確認
顧客データの保存場所、再委託先、越境移転有無	Cloud PF Type A は国内完結をうたうが、API 個別の詳細は 未確認 。 ⁴⁸
Web 検索/ブラウズ機能の有無と無効化可否	クローラーポリシーはあるが API 個別仕様は 未確認 。 ³⁰

確認項目	公開情報の状況
ファインチューニング、LoRA、専用埋め込み、専用 guard 提供可否	FAQ 上は現時点でカスタマイズ未対応。将来可能性あり。 ⁴⁹
閉域接続、VPC、専用線、SSO、SCIM、監査ログ連携	未確認
学習データの詳細プロビナンスと権利処理台帳	ベースソースは一部公開されるがサイト単位・データセット単位の詳細は 未確認 。 ⁵⁰
Embeddings API の内部モデル、次元数、日本語特化度、商用条件	未確認 。OSS v1/v2 はあるが商用制限あり。 ²¹
OCR/Vision を API で利用可能か、または OSS 前提か	未確認
出力物に関する補償、第三者権利侵害時の責任分担	未確認
導入支援体制、PoC 支援、評価テンプレート、専門家伴走の有無	一般的な問い合わせ窓口は確認できるが、個別体制は 未確認 。 ⁵¹

ソフトバンク/SB Intuitions への推奨質問は、次のように具体化するとよいです。

「Sarashina API で使用されるモデルバージョン、更新通知方式、コンテキスト長を提示してほしい」

「埋め込みモデルの次元数・保持期間・再学習不使用条項を契約に入れられるか」

「法務・知財用途での利用制限が API 契約上どう定義されるか」

「Web 参照機能がある場合、無効化と完全オフライン運用ができるか」

「Cloud PF Type A 上での専有環境、閉域接続、監査ログ、削除証跡の提供可否」

「OCR/Vision を組み込んだ知財文書処理構成を商用サポートできるか」

これらがクリアになれば、知財業務向けの採否判断はかなり前進します。 ⁵²

参考資料一覧

以下は、本調査で特に重要だった一次・準一次資料です。URL はご要望に合わせて列挙します。

- ソフトバンク 法人向け **Sarashina API**

<https://www.softbank.jp/business/service/ai/sarashina-api/> ⁵³

- ソフトバンク 企業・IR **Sarashina API 提供開始**

https://www.softbank.jp/corp/news/press/sbkk/2025/20251105_01/ ⁵⁴

- ソフトバンク 企業・IR **Cloud PF Type A 上で Sarashina 提供**

https://www.softbank.jp/corp/news/press/sbkk/2026/20260416_02/ ⁵⁵

- ソフトバンク 法人向け **Cloud PF Type A 提供開始**

<https://www.softbank.jp/biz/news/other/20260401/> ⁵⁶

- SB Intuitions プレス **2024年の公開モデル一覧**

https://www.sbintuitions.co.jp/news/press/20240614_01/ ⁵⁷

- SB Intuitions TECH BLOG **Sarashina1・2の仕様と学習データ**
<https://www.sbintuitions.co.jp/blog/entry/2024/06/26/115641> 58
- SB Intuitions TECH BLOG **Sarashina2-70Bの性能とトークナイザ**
<https://www.sbintuitions.co.jp/blog/entry/2024/08/21/144254> 11
- SB Intuitions プレス **Sarashina2-8x70B**
https://www.sbintuitions.co.jp/news/press/20241108_01/ 59
- Hugging Face **Sarashina2-8x70B**
<https://huggingface.co/sbintuitions/sarashina2-8x70b> 60
- Hugging Face **Sarashina2.2-1B**
<https://huggingface.co/sbintuitions/sarashina2.2-1b> 61
- Hugging Face **Sarashina2.2-1B-Instruct**
<https://huggingface.co/sbintuitions/sarashina2.2-1b-instruct-v0.1> 62
- SB Intuitions TECH BLOG **Sarashina2-Vision**
<https://www.sbintuitions.co.jp/blog/entry/2025/03/17/111659> 63
- Hugging Face **Sarashina2.2-Vision-3B**
<https://huggingface.co/sbintuitions/sarashina2.2-vision-3b> 64
- Hugging Face **Sarashina2.2-OCR**
<https://huggingface.co/sbintuitions/sarashina2.2-ocr> 65
- Hugging Face **Sarashina-Embedding-v1-1B**
<https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b> 66
- Hugging Face **Sarashina-Embedding-v2-1B**
<https://huggingface.co/sbintuitions/sarashina-embedding-v2-1b> 67
- SB Intuitions **情報セキュリティポリシー**
<https://www.sbintuitions.co.jp/security-policy/> 68
- SB Intuitions **個人情報保護方針**
<https://www.sbintuitions.co.jp/privacy-policy/> 69
- SB Intuitions **クローラーについて**
<https://www.sbintuitions.co.jp/bot/> 30
- 文化庁 **AIと著作権について**
<https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html> 70
- 文化庁 **AIと著作権に関する考え方について**
https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf 71

- 特許庁 AI関連発明

<https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/index.html> 72

- 特許庁 AI関連技術に関する特許審査の事例

https://www.jpo.go.jp/system/laws/rule/guideline/patent/ai_jirei.html 73

- J-PlatPat

<https://www.j-platpat.inpit.go.jp/> 74

1 33 36 37 62 <https://huggingface.co/sbintuitions/sarashina2.2-1b-instruct-v0.1>
<https://huggingface.co/sbintuitions/sarashina2.2-1b-instruct-v0.1>

2 13 26 <https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b/blob/main/LICENSE>
<https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b/blob/main/LICENSE>

3 8 57 https://www.sbintuitions.co.jp/news/press/20240614_01/
https://www.sbintuitions.co.jp/news/press/20240614_01/

4 22 27 44 45 47 49 52 53 <https://www.softbank.jp/business/service/ai/sarashina-api/>
<https://www.softbank.jp/business/service/ai/sarashina-api/>

5 9 41 42 46 54 https://www.softbank.jp/corp/news/press/sbkk/2025/20251105_01/
https://www.softbank.jp/corp/news/press/sbkk/2025/20251105_01/

6 35 74 <https://www.j-platpat.inpit.go.jp/>
<https://www.j-platpat.inpit.go.jp/>

7 34 <https://www.softbank.jp/biz/news/other/20260324/>
<https://www.softbank.jp/biz/news/other/20260324/>

10 23 29 50 58 <https://www.sbintuitions.co.jp/blog/entry/2024/06/26/115641>
<https://www.sbintuitions.co.jp/blog/entry/2024/06/26/115641>

11 24 25 39 43 <https://www.sbintuitions.co.jp/blog/entry/2024/08/21/144254>
<https://www.sbintuitions.co.jp/blog/entry/2024/08/21/144254>

12 59 https://www.sbintuitions.co.jp/news/press/20241108_01/
https://www.sbintuitions.co.jp/news/press/20241108_01/

14 15 <https://huggingface.co/sbintuitions/sarashina2.1-1b>
<https://huggingface.co/sbintuitions/sarashina2.1-1b>

16 61 <https://huggingface.co/sbintuitions/sarashina2.2-1b>
<https://huggingface.co/sbintuitions/sarashina2.2-1b>

17 63 <https://www.sbintuitions.co.jp/blog/entry/2025/03/17/111659>
<https://www.sbintuitions.co.jp/blog/entry/2025/03/17/111659>

18 31 64 <https://huggingface.co/sbintuitions/sarashina2.2-vision-3b>
<https://huggingface.co/sbintuitions/sarashina2.2-vision-3b>

19 20 38 40 65 <https://huggingface.co/sbintuitions/sarashina2.2-ocr>
<https://huggingface.co/sbintuitions/sarashina2.2-ocr>

21 32 66 <https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b>
<https://huggingface.co/sbintuitions/sarashina-embedding-v1-1b>

- 28 68 <https://www.sbintuitions.co.jp/security-policy/>
<https://www.sbintuitions.co.jp/security-policy/>
- 30 <https://www.sbintuitions.co.jp/bot/>
<https://www.sbintuitions.co.jp/bot/>
- 48 56 <https://www.softbank.jp/biz/news/other/20260401/>
<https://www.softbank.jp/biz/news/other/20260401/>
- 51 https://www.sbintuitions.co.jp/news/press/20251105_01/
https://www.sbintuitions.co.jp/news/press/20251105_01/
- 55 https://www.softbank.jp/corp/news/press/sbkk/2026/20260416_02/
https://www.softbank.jp/corp/news/press/sbkk/2026/20260416_02/
- 60 <https://huggingface.co/sbintuitions/sarashina2-8x70b>
<https://huggingface.co/sbintuitions/sarashina2-8x70b>
- 67 <https://huggingface.co/sbintuitions/sarashina-embedding-v2-1b>
<https://huggingface.co/sbintuitions/sarashina-embedding-v2-1b>
- 69 <https://www.sbintuitions.co.jp/privacy-policy/>
<https://www.sbintuitions.co.jp/privacy-policy/>
- 70 <https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html>
<https://www.bunka.go.jp/seisaku/chosakuken/aiandcopyright.html>
- 71 https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf
https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94037901_01.pdf
- 72 <https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/index.html>
<https://www.jpo.go.jp/system/patent/gaiyo/sesaku/ai/index.html>
- 73 https://www.jpo.go.jp/system/laws/rule/guideline/patent/ai_jirei.html
https://www.jpo.go.jp/system/laws/rule/guideline/patent/ai_jirei.html