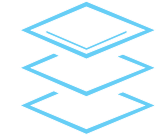


# 大規模案件も安心して任せられる、 AI開発・実装に欠かせないデータ提供パートナー



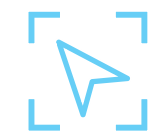
## 様々な開発シーンに活用可能な既製データセット

Nexdataは大規模な既製品データセットを保有しており、生成AI、ASR・TTSモデルなど幅広い用途に活用可能です。これによりお客様がAI開発におけるデータ収集の手間を省き、開発周期を大幅に短縮することができます。



## グローバルなデータ収集リソース

Nexdataは世界中にデータ収集リソースを広げています。これにより世界中のあらゆる地域から最適なデータを収集し、提供することが可能です。



## AI搭載した半自動アノテーションプラットフォーム

独自の専門的なプラットフォームは、画像・動画・音声・テキスト・点群などあらゆるモデルに対応可能です。AI自動識別と手動修正を組み合わせることでアノテーション効率を30%向上。お客様のローカル環境でのデータ処理も可能、データ資産が外部に出ることなくAI学習向けに転換可能です。



## 2万名多国籍アノテーターから迅速に最適なチーム編成

Nexdataにはグローバルなデータ処理施設と2万名を超える専門のアノテーターを擁し、お客様のニーズに最適なチームを迅速に編成可能です。日常的な教育および案件ごとの専門訓練・試験を通じ、生成AI・自動運転・フィジカルAIなど、お客様の期待を超える高品質データを納品しています。



## 大規模案件でも安心して任せられるプロジェクト管理

業界専門家が伴走し、柔軟なワークフローで段階的納品を実現。4段階の品質チェックを実施し、精度99.9%以上を確保しています。プロジェクトマネージャーは各国の大規模データ作成経験を有し、カスタマイズ要望に迅速かつ確に対応し、高品質データを納期通りにお届けします。

## 導入実績

Nexdataは創業以来10年以上にわたり、国内外5,000社以上のクライアント向けに、累計10,000件を超えるAIデータ提供プロジェクトを成功させ、世界のAIリーダー企業と最前線の研究機関を支えてきました。

今後もAI技術の進化を牽引する存在として、最先端のニーズに応えるデータソリューションを継続的に提供し、グローバルAI産業の発展を支えていきます。



# 日本最大級AI学習用データプロバイダー 学習データ調達なら、まずNexdataへ

## Nexdataが提供するデータソリューション



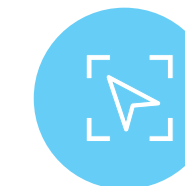
### 総計4.5PBのラベル付きデータセットを提供

Nexdataは、音声・画像・動画・テキストなど多様なモビリティで総計4.5PB規模の高品質データセットを保有・提供しています。AI開発に不可欠な大規模データをグローバル水準の品質管理体制のもと、お客様へ安定的にお届けします。



### AI学習データのカスタマイズ作成

大規模言語モデル(LLM)、自動運転、エンボディAIなど最先端分野に特化した学習データをお客様ごとの要件に合わせてカスタム開発します。豊富な開発実績を持つ専門チームが高品質かつスピーディーに大規模案件にも対応いたします。



### 独自のアノテーションプラットフォーム

Nexdata独自のAI識別機能を搭載したアノテーションプラットフォームは音声・画像・動画・テキストなど全モダリティに対応。これまでに1,000件以上のプロジェクトを支えてきた実績を活かし、大規模・高精度なデータラベリングを効率的に実現します。

## 人工知能データサービスプロバイダー



## 総計4.5PBデータセット、AI開発に即利用可能



自社著作権、データサンプルを無償提供可能



アノテーション付き、高精度な既製データ



カスタマイズ加工も可能、最短1日納品

### 100万時間音声認識・音声合成コーパス

- 50万時間日本語教師なし学習向けコーパス
- 633時間日本語自然会話音声コーパス
- 500時間LR分離日本語音声データセット
- 20万時間母語話者・訛り英語音声コーパス
- 100時間10種多言語音声合成コーパス
- 10.4時間日本語音声合成コーパス
- 500時間日本人話者英語LR分離対話音声データ
- 100時間日本語エンティティ含む読み上げ音声データ
- 5,000時間英語LR分離対話音声データ



### 800TB画像・動画データセット

- 20万ID、1,000万枚顔画像・動画データ
- 30万ID、100万人体画像・動画データ
- 2,500ID、100万点ジェスチャー画像データ
- 1,000万枚OCR画像データ



### LLM・マルチモーダルLLM向け

- 1PB教師なしテキストデータ
- 2.5億版權付き画像データ
- 800TB画像説明文データ
- 1PB動画説明文データ



### AI AGENT向け・エンボディドAI向け

- 20万点GUI AGENT向け操作画面データ
- 10万点以上の高精度3Dモデルデータ
- 20万件の「第一人称視点」人×物インタラクション動画
- 15万件のリアルタイム対話動画データ
- 15万件の3Dハンドジェスチャーデータ



### スマート運転向け

- 16,000ID運転手行動認識データ
- 7,000ID乗員行動認識データ
- 18種ジェスチャー識別データ
- 13カ国1,309時間走行行動動画データ



## 専門的な収集リソースで高品質なデータを生産

### 専門的な収録スタジオ



### エンボディドAIデータ収集センター



### グローバルリソースプール



## あらゆる業界・研究分野向けのAI学習データ収集が可能



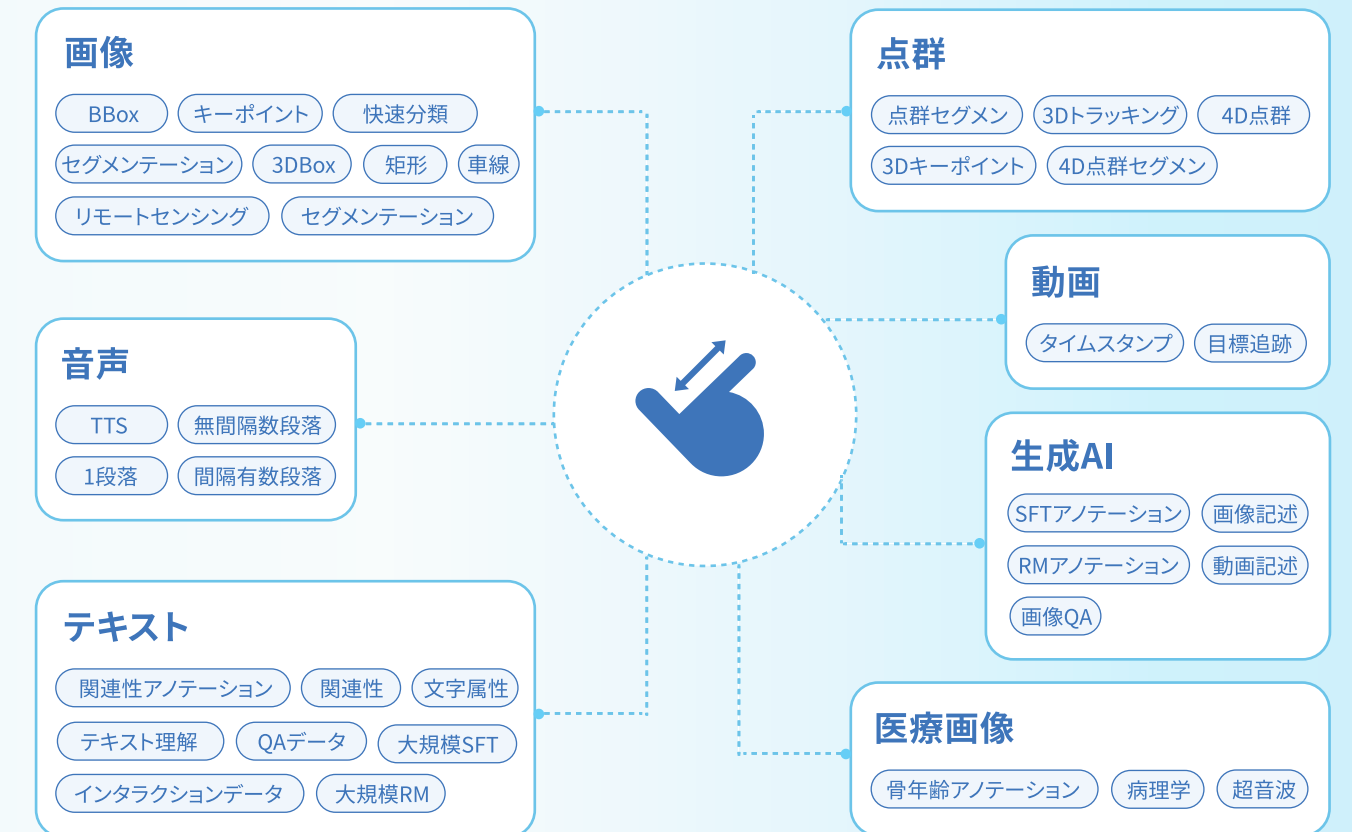
## さまざまなデータ収集に対応実績あり

音声	画像	動画	テキスト	点群
<ul style="list-style-type: none"> <li>読み上げ・自然会話・感情音声・声紋・騒音・WakeUp Word・会議音声などを収集可能</li> <li>日本語・英語・中国語など100種以上言語対応</li> <li>室内外、専門スタジオなど様々な収集環境・設備を整備</li> </ul>	<ul style="list-style-type: none"> <li>異なる人種・性別・年齢層・姿勢・動作・環境・設備における収集が可能</li> <li>製造業・建築業・自動車業・小売業・飲食業・医療業などの業界データ収集</li> <li>収集精度100%</li> </ul>	<ul style="list-style-type: none"> <li>物・人物・事件動画</li> <li>多感情・口の動作動画</li> <li>運動・スポーツ動画</li> <li>自動運転向け動画</li> <li>人型ロボット向け動画</li> </ul>	<ul style="list-style-type: none"> <li>教師なし学習向け</li> <li>教師あり学習向け</li> <li>SFT段階向け</li> <li>各種論文・教科書</li> <li>キャプションデータ</li> <li>COT推理推論データ</li> </ul>	<ul style="list-style-type: none"> <li>データ収集車</li> <li>自動運転に対応した点群データ収集可能</li> </ul>

## 独自アノテーションプラットフォーム納品まで一元管理

### 特徴:

- 35種のテンプレートを搭載、AI自動識別で作業効率が30%アップ
- 10,000件以上のデータプロジェクトによる検証済み
- 人的・ツール上にも安全管理を徹底、安心して依頼できる



## 様々なアノテーションに対応実績あり

音声	画像	動画	テキスト	点群
<ul style="list-style-type: none"> <li>音声分割</li> <li>クリーニング</li> <li>書き起こし</li> <li>騒音アノテーション</li> <li>話者アノテーション</li> </ul>	<ul style="list-style-type: none"> <li>Bbox</li> <li>ポリゴン</li> <li>セグメンテーション</li> <li>ライン</li> <li>3D Bbox</li> </ul>	<ul style="list-style-type: none"> <li>Bbox</li> <li>ポリゴン</li> <li>セグメンテーション</li> <li>行動認識</li> <li>行動追跡</li> </ul>	<ul style="list-style-type: none"> <li>クリーニング</li> <li>NER</li> <li>テキスト分類</li> <li>構文解析</li> <li>意図・感情認識</li> <li>QA・会話</li> </ul>	<ul style="list-style-type: none"> <li>3D物体検出</li> <li>3D物体追跡</li> <li>3D点群セグメンテーション</li> <li>4D点群アノテーション</li> </ul>