

外観検査AIを 使いこなす

撮像・学習データ・運用設計と、
最後に残る「検査員の判断」

*Mastering AI Visual Inspection:
Imaging, Training Data,
and the Judgment That Remains*

PUBLISHED BY

taiziii Research

Manufacturing DX Desk

本書の位置付け

本書は、製造業のDX担当者（事業部のDX推進部門、生産技術・製造・品質部門の部長・課長クラス）を読者に想定した、外観検査AIの実装に関する実務レポートである。「AIで検査を自動化しましょう」という掛け声ではなく、**どの照明とカメラで画像を取り、どんな学習データを作り込み、どのAIモデルを選び、どう人と分担しながら運用するか**という技術的・実務的な具体に踏み込む。

外観検査は、製品の品質を市場に出る前に最終確認する工程である。長年、検査員の目視と触感に頼ってきたが、人手不足と検査員の高齢化、判定のばらつきがその限界を露わにしている。深層学習（ディープラーニング）の進歩で、AI外観検査は中小製造業まで導入が広がってきた。

本書が一貫して主張するのは、外観検査AIの成否は「**高精度なAIモデルを買うこと**」ではなく「**画像が確実に撮れる照明・カメラの環境を作り、学習データを丁寧に整え、人とAIの分担を運用設計すること**」で決まるという点である。AIモデルの優劣は最後の段階の問題にすぎない。撮像が崩れていれば、どんなモデルでも欠陥は見えない。データが整っていなければ、どんなモデルでも学習できない。本書は、この技術以前の土台と、AIと共存する運用設計に重点を置く。なお本書は出荷前の検査工程を扱う — **不良の発生そのものを工程で減らす歩留まり改善は別テーマ**であり、検査は「最後の砦」として両者は補完関係にある。

ただし、土台を整えても消えないものが一つある。「**これは見逃せない**」と検査員が引っかかる、その判断だ。100%の精度を持たないAIを使うかぎり、境界の玉は必ず人のところへ戻ってくる。本書は撮像・データ・運用に踏み込んだうえで、最後にこの判断をどう引き出し、どう次へ残すかへ着地する。

本書は公開情報および当社の実プロジェクト知見（業界一般化・匿名化済み）に基づく。記載する数値（検査精度・効果等）は対象・撮像条件・運用により大きく変動するため、絶対値ではなく相対的な傾向として参照されたい。

目次

1 なぜいま外観検査AIなのか

目視の限界と、AIが補えるもの・補えないもの

2 教師あり学習と教師なし学習

2つの学習アプローチと、不良データが集まるかで選ぶ

3 撮像が9割 — 照明・カメラ・撮像環境

画像に欠陥が写らなければAIは何もできない

4 学習データの作り込み

画像収集・アノテーション・データ拡張・生成AIによる補完

5 過検出と見逃し — しきい値は誰が決めるのか

100%精度は不可能。だから、判断が人に残る

6 現場から — フロー図の一行の裏側

ある精密切削加工メーカーで見た、判断が書かれない理由

7 導入ロードマップ — 撮像を作り、判断を残す

PoCから始める順序、つまずき、最後に残る問い

外観検査AIの成否は、最新モデルを買うことではなく、撮像と学習データの土台を整え、人とAIの分担を運用設計することで決まる。そして土台を整えた先には、検査員の判断が残る。本書の要点を6つにまとめる。

1

目視検査は人手不足・検査員の高齢化・判定のばらつきという三重の限界に直面している

外観検査は長く検査員の目視と触感に依存してきた。しかし、検査員の高齢化と退職、集中力の限界による見逃し、人によるばらつき、24時間連続稼働の困難という限界が顕在化している。AIによる外観検査は、これらの限界を補い、品質と生産性の両立を実現する。経済産業省も導入を後押しするガイドブックを公開している。

2

学習アプローチは教師あり／教師なしの2系統。不良データが集まるかで選ぶ

教師あり学習は良品と不良品の両方を学習させ、不良の種類分類まで可能だが、不良画像を相当量集める必要があり、ラベル付け（アノテーション）の負担が大きい。教師なし学習（良品学習）は良品データのみを学習し「いつもと違う」を捉える。不良が稀／多様なケースに強い。実務では教師なしで始め、データが蓄積したら教師ありで種類分類へ、という二段構えも有効である。

3

外観検査AIは「照明×カメラ×AI」の掛け算。撮像が9割で、最新モデル選定は最後の問題

画像に欠陥が写っていないければ、どんな高性能AIでも判定できない。クロムめっきのような光沢面や凹凸の深い形状は撮像そのものが難しく、照明とカメラの設計がAI判定の上限を決める。導入で最初に投資すべきは、最新モデルではなく、欠陥が確実に写る撮像環境の構築である。

4

AIの実力は学習データの作り込みで決まる。アノテーションと不良データ不足が二大壁

AIは大量の画像から特徴を学ぶが、その画像とラベル（教師データ）を整えるのは人の仕事である。不良画像は発生頻度が低くて集まらない、欠陥領域のアノテーションは負担が大きい、という二大壁がある。データ拡張（回転・反転・明度変化）や、近年では生成AIによる疑似不良画像の生成で補完する手法も実用化しつつある。

5

しきい値は技術的な最適解ではない。過検出と見逃しのどちらをどこまで許すかという、人の判断である

ディープラーニングは統計的手法であり、原理的に100%の精度はありえない。判定を厳しくすれば良品まで不良と誤判定し（過検出）、緩めれば不良が市場に流出する（見逃し）。どこで線を引くかは、モデルが決められない。この部品のこの面なら、多少の過検出を飲んでも見逃しをゼロに寄せる — 熟練した検査員が部品ごとに下している重み付けが、そこにある。だから運用設計とは、「人をどこに残すか」の設計になる。

6

AIは判断しない。「これは見逃せない」を引き出し、条件付きルールとして構造化し、次の世代へ継ぐ

撮像もデータも運用も標準化できる。標準化できないものが、最後に一つ残る。検査員が一目見て何を疑うか、どこまでを不良と呼ぶか、という判断だ。それは検査基準書にも、しきい値にも、モデルの重みにも残らない。AIにこれを代行させるのではない。判断を引き出し、分類し、問い返し、次の担当へ渡す — その裏方に徹する。

01 なぜいま外観検査AIなのか

目視の限界と、AIが補えるもの・補えないもの

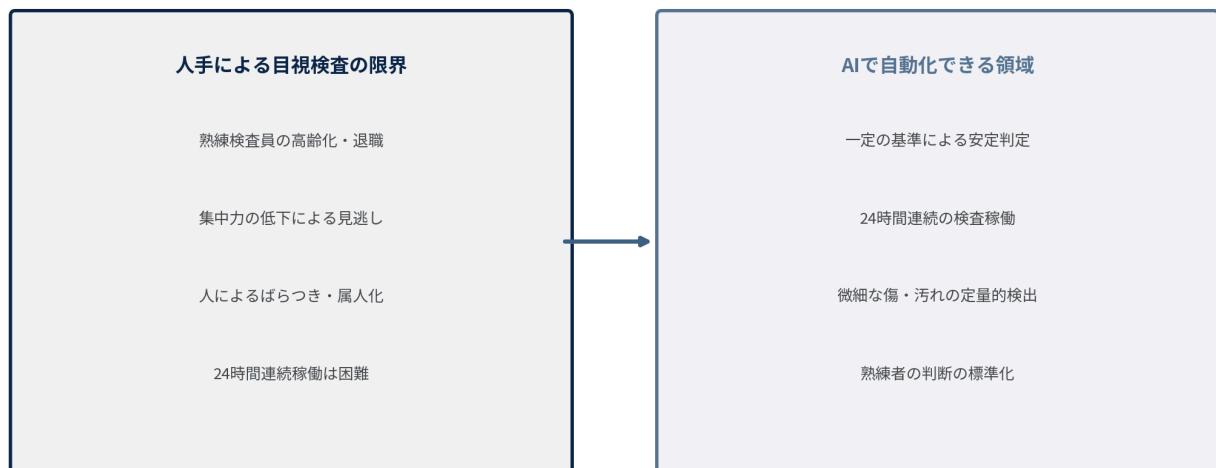
KEY FINDINGS

- 目視検査は熟練検査員の高齢化・退職、集中力低下による見逃し、属人化、24時間稼働の困難に直面している
- AIが補えるのは安定した一次判定・連続稼働・微細欠陥の定量検出まで。「これは見逃せない」という判断は人に残る
- 経済産業省は「中小企業向けAI導入ガイドブック 外観検査」を公開し、中小企業の導入を後押しする
- 機材費用は導入初心者向けで10万円～と例示されており、中小製造業でも始められる水準になった

1.1 目視検査が抱える三重の限界

外観検査は、製品の品質を市場に出る前に最終確認する重要な工程である。多くの製造現場では、長年、検査員が目で見手で触り、ベテランの勘で良品・不良品を判定してきた。しかし、その目視検査が、いま大きな限界に直面している。

なぜいま外観検査AIなのか



目視に頼り続けるか、AIを"道具として"使いこなすかが、品質保証と生産性の分かれ目になる

出所：taiziii Research 作成

第一に、**熟練検査員の高齢化と退職**である。微細な欠陥を見抜く眼を持つベテランが引退すると、その属人的能力は失われる。第二に、**集中力低下による見逃し**。人間の集中力は長時間は続かず、疲労や体調で精度がぶれる。第三に、**人によるばらつき・属人化**。誰が検査するかで判定が変わるため、品質保証のレベルが安定しない。加えて、人は24時間連続では働けない。

1.2 AIで補える領域・補えない領域

これに対し、AIによる外観検査は、目視の限界を補える領域を持つ。一定の基準による安定した一次判定、24時間連続稼働、微細な傷・汚れの定量的検出 — いずれも目視が苦手とする領域である。集中力の低下も疲労もないため、判定は安定する。

だが、AIが目視のすべてを代替できるわけではない。AIは画像から学んだパターンに基づいて判定するため、学習していない新種の欠陥には弱い。また、後述するように100%の精度はあり得ない。

補えない領域は、もう一段深いところにもある。**検査員が一目見て「この陰りは見逃せない」と引っかかる、その判断**だ。どこを疑うか。寸法内の打痕を通すか止めるか。この部品なら過検出を飲んででも見逃しを避けるべきか。こうした見極めは検査基準書にもしきい値にも書かれていない。**本書はこの判断をAIで均す（標準化する）のではなく、どう引き出し、どう次へ残すかを最後の主題にする**。AIは「人の代わり」ではなく、「人の負担を変え、判定を安定させる道具」である。本書はこの前提で論を進める。

1.3 経済産業省の後押しと中小企業の追い風

外観検査AIの普及には、もう一つ重要な背景がある。**経済産業省が「中小企業向けAI導入ガイドブック」**として**外観検査を主要テーマに据え、導入手順を公開している**ことだ。需要予測と並んで、外観検査は「中小企業がAIを活用する2大領域」として位置づけられている。同ガイドブックは、不良品が十分出ている場合（不良品あり）と、ほとんど出ていない場合（良品のみ）の2パターンに分けて、業務フロー、画像データの撮影方法、機材費用、効率化効果まで具体的に整理する。機材費用は導入初心者向けで10万円～と例示されており、中小製造業でも始められる水準である。本書も、随所でこの公的ガイドブックの枠組みに沿って論じる。

本書の立場：AIは判定の下働き。土台を整えた先に、検査員の判断が残る

外観検査AIに過剰な期待を抱くと、必ず失望する。AIは魔法の箱ではなく、撮像・データ・運用という土台の上で初めて機能する道具である。本書は、AIモデルの優劣ではなく、その土台の作り方と、人とAIの分担の設計に重点を置く。

そして、土台を整えるほど、残るものがはっきりしてくる。AIが担うのは安定した一次判定という下働きであり、境界の玉をどう裁くかという判断は人の側に残る。本書は最後に、その判断をどう継ぐかへ着地する。

02 教師あり学習と教師なし学習 2つの学習アプローチを使い分ける

KEY FINDINGS

- 外観検査AIには教師あり学習と教師なし学習の2系統がある。どちらを選ぶかで構築方法が大きく変わる
- 教師あり学習は不良の種類分類まで可能だが、不良画像とラベル付けが大量に必要
- 教師なし学習（良品学習）は良品データのみで「いつもと違う」を検知。不良データが集まらないケースに強い
- 選択基準は「不良品データが十分集まるか」。経産省ガイドブックも不良品あり／良品のみで章を分ける

2.1 2つの学習アプローチ

外観検査AIで使われる機械学習には、大きく2つのアプローチがある。**教師あり学習**と**教師なし学習**である。経済産業省のAI導入ガイドブックも、不良品が十分ある場合（不良品あり）と、ほとんど出ていない場合（良品のみ）の2パターンに章を分けて整理しているが、これは実は教師あり／教師なしの選択と対応している。

2つの学習アプローチ：どちらを採るかは「不良データが集まるか」で決まる

教師あり学習	教師なし学習（良品学習）
良品と不良品の両方を学習	良品データのみを学習
得意なこと 不良の種類（傷／打痕／さび等）まで分類できる	得意なこと 未知の不良パターンも「いつもと違う」として検知
必要なもの 良品データ＋十分な量の不良画像＋ラベル付け	必要なもの 十分な量の良品画像のみ（不良データ不要）
壁 不良品が十分集まらない・アノテーションが大変	壁 不良の「種類分類」は基本的にできない
向く対象 不良パターンが既知で、種類分類が必要なケース	向く対象 不良が稀／多様で集めにくいケース、初期導入

実務では、まず教師なしで「いつもと違う」を捉え、不良が集まったら教師ありで種類分類へ、という二段構えも有効

出所：機械学習の一般的な考え方および経済産業省「AI導入ガイドブック 外観検査」に基づき taiziii Research 作成

2.2 教師あり学習 — 不良の種類まで分類できる代わりに

教師あり学習は、良品と不良品の両方の画像を集め、それぞれに「これは良品」「これは傷の不良」「これは打痕の不良」といったラベル（教師）を人が付けて学習させる方法である。AIはこのラベル付き

画像から特徴を学び、未知の画像に対しても良品／不良品の判定、さらには**不良の種類まで分類**できるようになる。生産改善で「どの種類の不良が増えているか」を追いたいときは、この分類が効いてくる。

ただし代償も大きい。第一に、**不良品の画像を相当量集める必要がある**。良品ばかりが出る工程では、不良画像が数ヶ月かけても十分に集まらないことがある。第二に、**アノテーション（ラベル付け）が大きな負担になる**。単に「不良」と分類するだけならまだしも、欠陥の領域を画像上で一つひとつ囲ったり、ピクセル単位で塗りつぶしたりする必要があるケースでは、その作業量は膨大になる。

2.3 教師なし学習 — 「いつもと違う」を捉える

教師なし学習（外観検査では「良品学習」とも呼ばれる）は、別のアプローチをとる。**良品データのみを学習**させて「正常な状態とはこういうものだ」をAIに覚えさせ、それから外れた画像を「異常」として検知する。不良品のデータは不要であり、ラベル付けも基本的に要らない。

この発想は、予知保全で使われる異常検知と同じ考え方である。**故障や不良のデータは集めにくい**が、**正常データなら大量に手に入る**という製造業の現実、最も合致したアプローチといえる。さらに、学習していない未知の不良パターンも「いつもと違う」として捉えられる柔軟性がある。一方、弱点は**不良の種類分類は基本的にできない**こと。「何かが違う」とは分かっても「どの種類の不良か」は判定対象外である。

2.4 二段構えで使う

実務では、両方を排他的に選ぶ必要はない。**まず教師なし学習で運用を始め、不良データが蓄積したら教師あり学習で種類分類を上乗せする**、という二段構えが現実的で強力である。初期は不良データが集まらなくても運用を開始でき、データの蓄積と共にAIの能力を高めていける。経済産業省ガイドブックの2パターンも、企業の状況に合わせてどちらから始めるかを選ぶ枠組みとして読める。

ここで一つ、後の章への伏線を置いておきたい。教師あり学習が学ぶのは、人が付けたラベルである。そのラベルは、**検査員が「これは不良だ」と判断した結果**であって、判断そのものではない。なぜそれを不良と呼んだのか — その理由は、ラベルのどこにも書かれていない（第5章）。

03 撮像が9割 — 照明・カメラ・撮像環境

画像に欠陥が写らなければAIは何もできない

KEY FINDINGS

- 外観検査AIは「照明×カメラ×AIモデル」の掛け算。撮像が崩れるとAIモデルは機能しない
- 画像に欠陥が写っていないければ、どんな高性能AIでも判定できない。撮像が上限を決める
- 光沢面・凹凸の深い形状・透明体など、撮像そのものが難しい対象は照明とカメラの設計が必要
- AI導入で最初に投資すべきは、最新モデルではなく欠陥が確実に写る撮像環境の構築である

3.1 「掛け算」で決まる

外観検査AIの結果は、3つの要素の**掛け算**で決まる。①照明、②カメラ・レンズ、③AIモデル。どれか一つでも崩れると全体が立ち行かなくなる、ということだ。

外観検査AIは「掛け算」で決まる — 撮像が9割



画像に欠陥が写っていないければ、どんな高性能AIでも判定できない。①②が崩れると③は機能しない

AI導入で最初に投資すべきは、最新モデルでなく、欠陥が確実に写る照明とカメラの環境設計である

出所：taiziii Research 作成

そして、この3要素の中で**最初の2つ（撮像）が9割を占める**。なぜか。**画像に欠陥が写っていないければ、どんな高性能AIでも判定できない**からだ。AIは画像から欠陥の特徴を学び識別するが、その画像に欠陥の情報が含まれていなければ、学ぶことも識別することもできない。AIモデルの精度は、撮像が決めた「画像の質」という上限を超えることはない。

3.2 撮像が難しい対象は技術的挑戦

撮像が容易なら問題は少ない。難しいのは、**撮像そのものが技術的挑戦になる対象**である。代表例がいくつかある。**光沢面**（鏡面仕上げの金属、めっき加工品）は、光の反射で映り込みが入り、欠陥と紛れ

る。**凹凸の深い形状**は、影が出て陰になった部分の欠陥が見えない。**透明体**（ガラス、樹脂）は、そもそも画像に映る特徴が薄い。**暗い色や黒色の対象**は、輝度差が出にくく欠陥のコントラストが弱い。

こうした難しい対象には、専用の撮像設計が要る。光の当て方（拡散光、同軸落射、バックライト、リング照明など）、波長の選択（可視光、近赤外、紫外）、カメラの画素数とフレームレート、レンズの選定。これらを対象ごとにチューニングすることで、初めて欠陥が「写る」状態が作られる。光沢面や凹凸の深い形状は、多くの検査機器メーカーから「撮像できない」と断られることさえある領域だが、逆に言えば、撮像を作り込めればAI以前の勝負がついている。

3.3 ルールベース画像処理との使い分け

ここで補足しておきたいのが、**すべてをAIで解く必要はない**という点である。寸法測定や決まった位置のキズ検査など、ルールが明確で条件が安定している検査には、従来のルールベース画像処理のほうが高速・低コスト・説明可能で適している。AIが本領を発揮するのは、ルール化が難しい外観検査 — 形状や色合いが微妙に変わる、ばらつきが大きい、官能検査に近い領域である。**AIとルールベースは対立する選択肢ではなく、検査対象に応じて使い分けるもの**と捉えるのが現実的である。

最初に投資すべきは「最新モデル」ではなく「撮像環境」

AI外観検査の導入で予算が限られているとき、「最新のディープラーニングモデルを買う」より「欠陥が確実に写る照明・カメラの環境を作る」ほうが、結果として遥かに大きなリターンを生む。撮像環境への投資は、AIモデルの能力を引き出すための土台投資であり、これを軽視するとプロジェクト全体が立ち行かなくなる。

04 学習データの作り込み

画像収集・アノテーション・データ拡張・生成AIによる補完

KEY FINDINGS

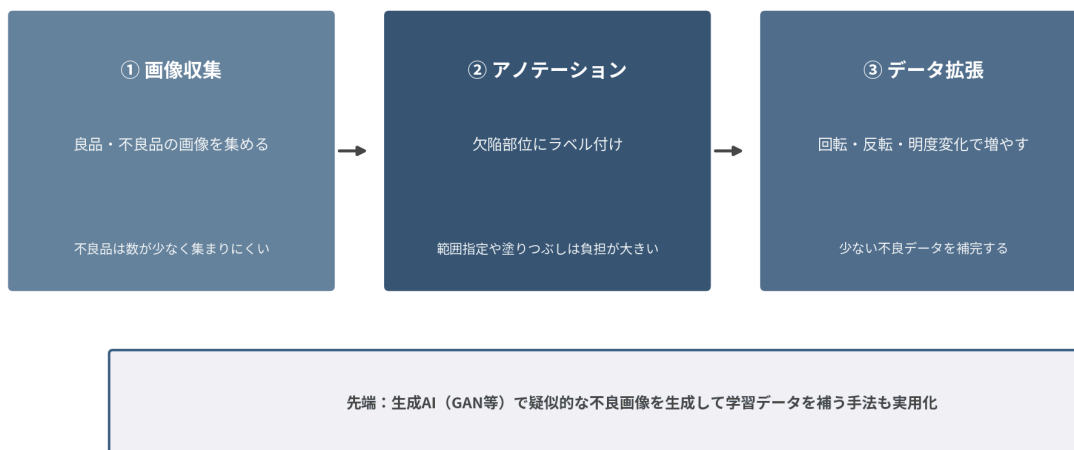
- AIの実力は学習データの質と量で決まる。データの作り込みが現場では最大の工数を占める
- 不良画像は発生頻度が低いため、収集自体に時間がかかる。教師なし学習の選択もこれが背景にある
- アノテーション（欠陥領域のラベル付け）は負担が大きい。塗りつぶし型は特に重い作業になる
- データ拡張（回転・反転・明度変化）や生成AIによる疑似不良画像で、データ不足を補う手法がある

4.1 データの作り込みが「AIの実力」を決める

AIは大量の画像から欠陥の特徴を自動で学ぶ。だが、その「大量の画像」とラベルを整えるのは、まぎれもなく人の仕事である。AIモデルの選定やパラメータ調整は、外部のソリューションや専門家の力を借りれば一部を任せられる。しかし、自社製品の良品・不良品画像を集めること、そこに正しいラベルを付けること、これは現場でしかできない。**外観検査AIプロジェクトで最も多くの工数を占めるのは、AIモデル開発ではなく、学習データの作り込み**であることが多い。

学習データの作り込みが「AIの実力」を決める

AIモデルの選定より、学習用データの収集とラベル付けが、現場では最大の工数を占める



出所：taiziii Research 作成

4.2 二大壁 — 不良画像の希少性とアノテーションの負担

データの作り込みには、2つの大きな壁がある。第一が**不良画像の希少性**。歩留まりが高い工程ほど不良が出ず、学習に必要な不良画像が集まらない。数ヶ月待っても十分なサンプル数に達しないことは珍しくない。故障や不良のデータは集めにくい — これはAIを製造業に持ち込むとき、たいていの現場で共通してぶつかる壁だ。だから第2章で述べたとおり、教師なし学習（良品データのみで学ぶ）が現実的な選択肢になる。

第二が**アノテーション（ラベル付け）の負担**。単純な良品／不良品の二分類ならまだしも、欠陥の領域を一つひとつ囲ったり（バウンディングボックス）、ピクセル単位で塗りつぶしたり（セグメンテーション）する作業は、慣れた人でも一枚に数分かかる。何千枚もの画像を処理する負担は大きく、これがAI導入のボトルネックになる。

ただし、この負担の正体は作業量だけではない。担当者が変わるとラベルが揺れる、境界の一枚で手が止まる、複数人で付けると基準が合わない — アノテーションの現場でよく起きるこれらは、「**どこまでを不良と呼ぶか**」という判定を、**一枚ごとにやり直している**ということだ。その判定基準が言葉になっていなければ、枚数をこなしても揺れは消えない。

4.3 補う技術 — データ拡張と生成AI

この2大壁を技術で補う手法も発達している。代表的なのが**データ拡張（Data Augmentation）**。少ない元画像を、回転・反転・明度変化・ノイズ追加といった変換でバリエーションを増やし、学習データを実質的に水増しする。1枚の不良画像を10枚相当に拡張する、といった効果を得られる。

さらに先端では、**生成AI（GAN：敵対的生成ネットワークなど）**で疑似的な不良画像を生成し、学習データを補う手法が実用化しつつある。実際の不良に近い特徴を持つ画像を人工的に作り出すことで、希少な不良データの不足を補える。完全には実不良の代わりにならないが、初期の学習を立ち上げる足がかりにはなる。これらはAIモデルの工夫というより、**データの土台を補強する技術**だ。本書が言う「データの作り込みが9割」を、技術の面から後押しする。

05 過検出と見逃し — しきい値は誰が決めるのか

100%精度は不可能。だから、判断が人に残る

KEY FINDINGS

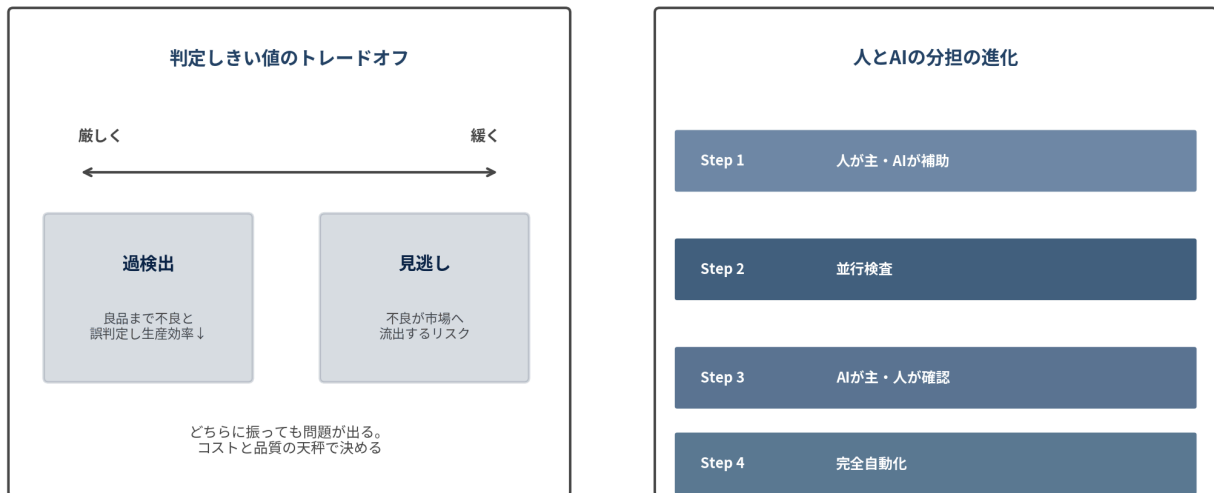
- ディープラーニングは統計的手法。原理的に100%精度は実現できないことを前提に運用設計する
- 判定しきい値を厳しくすれば過検出（良品を不良と誤判定）、緩めれば見逃しが増えるトレードオフがある
- しきい値は技術的な最適解ではなく、どちらの損をどこまで飲むかという人のトレードオフ判断である
- 判断には5つの粒度がある。検査基準書に書けるのは上の2つだけで、しきい値はいちばん下の層にある
- 人とAIの分担とは、人を残すことではなく、残った判断を記録できる形にすることである

5.1 100%精度はあり得ない

AIによる外観検査について、最初に正しく理解しておくべきことがある。**ディープラーニングは統計的な手法であり、原理的に100%の精度は実現できない。**AIが出すのは「これは99%の確率で不良品である」といった確率値であり、100個に1個は間違える可能性がある。不良品をすべて見つけるAIは、しくみ上あり得ない。

この前提を認めたうえで、外観検査AIの運用は「では、どうやってミスを管理するか」という問いに集約される。問われるのは2つ。**(1) どこまでをAIに任せるか、そして(2) 人をどこに残すか。**ここから先は技術ではなく、運用の設計になる。

100%精度は原理的に不可能。だから「人とAIの分担」を設計する



トレードオフを認め、人を残しつつ段階的に自動化を進めることが現実的な運用設計になる

出所：taiziii Research 作成

5.2 過検出と見逃しのトレードオフ

AIは確率値を出すため、「確率が何%以上なら不良と判定するか」というしきい値を決める必要がある。このしきい値設定が、運用上の重要な意思決定になる。

しきい値を**厳しく**すると（低い不良確率でも不良と判定する）、不良の見逃しは減るが、**良品まで不良と誤判定する過検出**が増える。過検出が増えると、再検査の手間が増え、廃棄や手直しのコストがかさみ、生産効率が落ちる。

しきい値を**緩く**すると、過検出は減るが、**不良の見逃し**が増える。見逃した不良は市場に流出し、クレーム、リコール、信用失墜につながりうる。安全関連の重要部品では絶対に避けたい事態である。

つまり、過検出と見逃しは**トレードオフの関係**にあり、どちらかをゼロにすることはできない。最適なしきい値は、対象部品の重要度、不良流出のリスク、過検出のコストを天秤にかけて設定するほかない。ここでよく使われるのが「技術的な最適解ではなく、経営判断だ」という言い方である。だが、その一言で片づけると、いちばん大事なものを取り落とす。

「傷は長さ0.1mm以上で不良」という閾値は、知識（What to do）として書ける。だが「この部品のこの面なら、多少の過検出を飲んででも見逃しをゼロに寄せる」という重み付けは、書けない。それは熟練した検査員が部品ごとに、ほとんど無意識に下しているトレードオフ判断（How to decide）である。**しきい値の議論が「経営判断」という言葉で止まってしまうのは、この判断が、誰の口からも言葉として出てこないからだ。**

5.3 判断には5つの粒度がある

判断を残そう、とだけ言っても話は前に進まない。判断をひとかたまりに扱っているからだ。実際には、判断は次の5つの粒度に分解できる。そして粒度ごとに、残せるかがどうかがまるで違う。

粒度	性質	外観検査での例	文書化
明示ルール	数値閾値・Yes/No	傷は長さ0.1mm以上で不良	できる
条件付きルール	If-Then-Else	摺動面かつ深さのある打痕なら、寸法内でも止める	できる
ヒューリスティック	経験則・傾向	このロットは端に陰りが出やすい	難しい
パターン認識	言語化しにくい直感	見た瞬間に、どこを疑うかが決まっている	できない
トレードオフ判断	二律背反の重み付け	過検出を飲んでも、見逃しをゼロに寄せるか	できない

検査基準書に書き切れるのは、上の2つだけ。下の3つは、本人に聞いても「見れば分かる」としか返ってこないことが多い。

しきい値は、いちばん下のトレードオフ判断である。それを、いちばん上の道具（数値の閾値）で決めている。しきい値の会議がいつまでも決着しないのは、担当者の力量の問題ではない。道具と対象の階層がずれているからだ。

そして、検査員が製品を見た瞬間に「どこを疑うか」を決めているのは、下から2番目のパターン認識にあたる。これがどう現場に現れるかは、第6章で具体を示す。

しきい値を振り直しても、精度が上がらなかった現場へ

撮像を作り直し、データを増やし、しきい値を何度も振り直した。それでも過検出は減らず、現場は結局AIの判定を目で見直している。

これは、取り組みが甘かったからではない。しきい値は精度の問題ではなく、どこまでを不良と呼ぶかという判断の問題だからだ。その判断が言葉になっていないかぎり、どのモデルに乗り換えても、同じ場所で止まる。

5.4 人とAIの分担を段階的に進化させる

100%精度がない以上、人をどこかに残す必要がある。だが、最初からすべてをAIに任せる必要も、いつまでも人が全件確認する必要もない。**人とAIの分担を、段階的に進化させる**のが現実的な定石である。

初期段階（Step 1）は**人が主・AIが補助**。判定は人が行い、AIは候補を提示するだけにとどめる。次の段階（Step 2）は**並行検査**。人とAIが両方検査して結果を突き合わせ、AIの判定精度を客観的に把握する。AIの精度に信頼が積み上がったなら（Step 3）、**AIが主・人が確認**に切り替える。AIの判定結果のうち、確率が境界線上の事例だけを人が確認する。最終的に（Step 4）、明らかな良品・不良品は**完全自動化**し、人は境界事例に集中する。

この段階を急いではいけない。Step 1からStep 2、Step 2からStep 3へ進む判断は、十分な事例検証を経て行う。**AIへの過信は、過小評価と同じくらい危険**である。

ここで注意したいのは、最終段階で人に残るのは「例外対応」という後始末ではない、ということだ。**AIが迷う境界事例こそ、判断が要る場所である**。AIが「いつもと違う」を一万件出したとして、そのうち本当に危険な数件を選び分けてきたのは検査員の判断だった。だとすれば、そこで**なぜそう判定したのかを**

理由ごと記録することが、言葉にならなかった判断をデータへ翻訳する、ほとんど唯一の入口になる。人とAIの分担とは、人を残すことではない。**残った判断を、記録できる形にすることだ。**

運用の型と、判断の型

運用の型は、誰がやっても同じ結果になる手順の型である。ここではAIは効率化装置として働き、自動化やRPAが効く。**判断の型**は、例外に直面したとき熟練者がどう判断するかの型である。

多くの企業は、運用の型を作ったところで止まる。外観検査AIの導入が「精度は出た。だが結局、最後はベテランが見ている」で止まるのは、判断の型に手をつけていないからだ。前者を自動化しても、後者は一ミリも継がれない。

5.5 引き出し、構造化し、継ぐ

では、判断の型はどう残すのか。順序は3つである。

引き出す。境界事例を判定したその場で、「どこを見て、なぜそう決めたか」を聞く。あとから思い出してもらうのでは出てこない。判定の瞬間に問うから、言葉になる。

構造化する。出てきた判断を、条件付きルールの形へ落とす。「摺動面／深さのある打痕／寸法内」ならこう裁く、というように。このとき**帰属を保つ**ことが要になる。誰の、どの条件下での判断か。ここを消して平均すると、判断は無害な一般論に薄まって使えなくなる。**判断を平均しない。一本の正解に統合しない。例外は例外として分けて残す。**

継ぐ。構造化された判断は、次の担当が同じ場面に立ったとき、比較の材料になる。「前任者はこの条件でこう裁いた」。それを見たうえで、どう判断するかは本人が決める。

この一連の作業で、**AIは判断しない**。判断の場面を捉え、問いを投げ、返ってきた言葉を分類し、矛盾や抜けを問い返す。やるのはそこまでである。判定を代行させれば、判断は今度こそ誰の頭からも消える。

検査員が一人抜けたあと、何が起きるか

ベテランの検査員が抜けた現場に起きるのは、不良率がわずかに動くといった穏やかな変化ではない。まず、**見逃しが出る**。基準書には載っていない「疑うべき場所」を、誰も見なくなる。次に、**判定がぶれる**。人によって線引きが変わり、品質保証のレベルそのものが定まらない。やがて、**安全側に倒れる**。迷った玉をすべて不良側に落とすので、過検出が積み上がり、歩留まりと工数が悪化する。そして最後に、AIも育たなくなる。**なぜ不良と呼んだのかが残っていないので、再学習に使える教師データが作れない。**

06 現場から — フロー図の — 行の裏側

ある精密切削加工メーカーで見た、判断が書
られない理由

KEY FINDINGS

- 業務フローを書き出してもらおうと一本の線になり、各工程は数行で終わる。検査は、たった一語だった
- 書けなかったのは知識が足りないからではない。判断が体に染み込みすぎて、書くほどのことだと思われていなかった
- 職人は図面を見た瞬間に段取りパターンを頭で展開する。検査員が「どこを疑うか」を決めるのと同じ層の判断である
- 基準書（管理側の認識）と現場の実際は食い違う。しかも双方とも、そのズレに気づいていない
- 「傷」「打痕」といった基本用語ですら、書く側と見る側で指す範囲がずれていることがある

ここまでの章で、外観検査AIの土台（撮像・データ・運用）と、その先に残る判断について述べてきた。本章では、その「残る判断」が現場でどう現れるかを、当社が実際に支援した案件から示す。外観検査そのものの案件ではない。ある精密切削加工メーカーで、基幹システムと製造現場のデジタルツインを構築した案件である。だが、そこで起きたことは、外観検査AIを入れようとする現場でそのまま起きる。

6.1 数行で終わった業務フロー

最初に、業務フローを書き出してもらった。返ってきたのは「受注 → 設計 → NCプログラム作成 → 加工 → 検査 → 出荷」という一本の線で、各工程はわずか数行のメモで終わっていた。**検査は、一語だった。**

これを「書けていない」と受け取ったのは、支援した側の浅さだった。書けなかったのは、知識が足りないからではない。逆である。**判断が体に染み込みすぎていて、本人たちが「書くほどのこと」だと思っていなかった。**フロー図は数行で書ける。その数行の裏に、何十年分の判断が詰まっていた。

6.2 一行の裏にあった判断

その一行の裏で、熟練が何をしていたか。三つ書く。

見た瞬間に、頭の中で全部を展開している。ワークには複数の面があり、それぞれ違う角度から削る。どの面から加工するかで、次の段取りのための固定し直し方も、基準面の取り方も、使う工具も変わる。面の数と角度の組み合わせだけ段取りパターンがあり、職人は図面を一枚見た瞬間に、そのパターンを頭の中で展開している。本人に聞いても「見れば分かる」としか返ってこない。

検査員が製品を見た瞬間に「どこを疑うか」を決めているのも、同じ層の判断だ（第5章の表でいうパターン認識）。言葉にならないまま、退職とともに消える。

基準書と現場は、食い違っている。管理層が共有した前提は明快だった。工具は使うたびに組み立て、使い終わったら分解して返却する。システムもその前提で設計されかけた。ところが現場の実態は違った。**使い終えた工具はあえてバラさず保管し、次に同じ構成が要るときに再利用していた。**段取り時間を削る

ために現場が編み出した知恵である。管理層の認識と現場の実際がずれていて、しかも双方とも気づいていなかった。前提のまま作ってれば、動きはするが現場と噛み合わないシステムを納めていたはずだ。検査でも同じことが起きる。検査基準書に書かれた見るべき場所と、検査員が実際に見ている場所は、同じとは限らない。そこを確かめずにアノテーションの基準を決めれば、基準書どおりで現場と合わないラベルが積み上がる。

言葉そのものが、共有されていない。「プログラム」「ワーク」といった基本用語が現場で独自の意味を持ち、書く側と動かす側で指す範囲までずれていた。密度の濃い現場ほど、用語は内部で進化して外部の共通語から離れる。同じ単語で会話しているのに、互いに違うものを指している。

「傷」「打痕」「シミ」も、外から思うほど共有された言葉ではない。複数人でアノテーションするとラベルが揃わない。その原因の一部は、作業量でもモデルでもなく、ここにある。

ツールが無いから難しいのではない

この案件で分かったのは、単純だが厄介な事実だった。**難しいのはツールが足りないからではなく、現場の判断がまだ言葉になっていないからだ。**ここが埋まらないかぎり、どのツールに乗り換えても、同じ場所で止まる。

外観検査AIも例外ではない。撮像を作り、データを整え、しきい値を振り直しても、「検査」「判定」という一語に畳まれた判断が言葉にならなければ、そこで止まる。フロー図に一語で書かれているものほど、疑ったほうがいい。

6.3 では、どこから手をつけるか

この現場で最初にやったのは、新しいツールを選ぶことではなく、判断を一つずつ言葉にしていく作業だった。順序は第5章で述べたとおりである。判断が起きるその場で「なぜそう決めたか」を引き出し、帰属を保ったまま条件付きルールへ構造化し、次の担当が参照できる形にする。地味で、時間がかかる。だが、ここを飛ばして組んだ仕組みは、動いても使われない。

外観検査AIでいえば、それは「AIが迷った一枚を、ベテランがなぜそう裁いたか」を記録するところから始まる。次章で、その進め方を導入口ードマップの中に位置づける。

07 導入ロードマップ — 撮像を作り、判断を残す

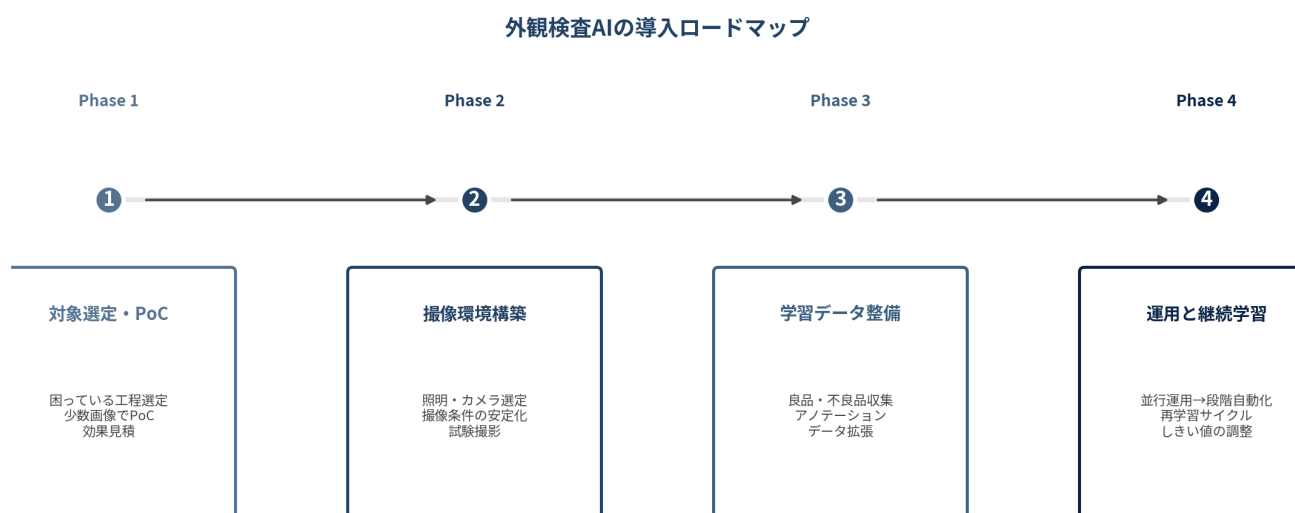
PoCから始める順序、つまり、最後に残る問い

KEY FINDINGS

- 導入は「対象選定・PoC→撮像環境構築→学習データ整備→運用と継続学習」の4段階で進める
- 最初は1工程・1製品でPoC（概念実証）を行い、撮像と精度の見通しを立ててから本格展開する
- 経済産業省「AI導入ガイドブック 外観検査」が公的な導入手順として参照できる
- 典型的なつまずきは、撮像軽視、データ不足の軽視、しきい値固定運用、人を完全に外した過信
- 撮像も、データも、運用も標準化できる。標準化できないのは、検査員の判断である

7.1 4段階のロードマップ

外観検査AIの導入は、思いつきの試行錯誤ではなく、段階を踏んで進める。



経済産業省「中小企業向けAI導入ガイドブック 外観検査」も同様の段階で導入を整理している

出所：経済産業省「AI導入ガイドブック 外観検査」の枠組みに基づき taiziii Research 作成

Phase 1：対象選定・PoC。困っている検査工程・製品を選び、少数の画像で概念実証（PoC）を行う。AIで解けそうか、効果がどれくらい見込めるかを試算する。**Phase 2：撮像環境構築。**本書第3章のとおり、撮像が9割。照明とカメラを選定し、撮像条件を安定化させる。試験撮影を繰り返して画質を作り込む。**Phase 3：学習データ整備。**本書第4章のとおり、データの作り込みがAIの実力を決める。良品・不良品の画像収集、アノテーション、必要ならデータ拡張・生成AIによる補完を行う。**Phase 4：運用と継続学習。**本書第5章のとおり、人とAIの並行運用から始め、段階的に自動化を進める。定期的な再学習としきい値の調整で精度を維持する。経済産業省の「AI導入ガイドブック 外観検査」も、同様の段階で導入を整理している。公的な手順書として参照すると、検討の漏れを防げる。

7.2 定着のコツ — 「育て続ける」

外観検査AIは、導入したら終わりではない。レンズや照明の汚れ、機械の経年劣化、環境光の変化、製品仕様の変更 — 現場の小さな変化が、検査精度に効いてくる。**定期的な点検・清掃、必要に応じた再学習、しきい値の見直し**が、長期運用の鍵になる。設置して放置すれば、判定は少しずつずれていく。崩れた分を点検と再学習で戻し続けて、はじめて精度は保たれる。

典型的な4つのつまずき

① **撮像軽視**：AIモデル選定に予算を集中し、照明・カメラへの投資をケチる。結果、欠陥が画像に写らずAIは何もできない。最初に投資すべきは撮像環境である（第3章）。② **データ不足の軽視**：不良画像が集まらないのに教師あり学習を強行し、運用開始が大きく遅れる。教師なし学習やデータ拡張・生成AIによる補完を組み合わせる（第2章・第4章）。③ **しきい値固定運用**：一度設定したしきい値を見直さず、過検出や見逃しが累積する。定期的な調整が要る（第5章）。④ **人を完全に外した過信**：100%精度にならないAIに全件判定を任せ、重大な見逃しを招く。人とAIの分担を段階的に進化させる（第5章）。

7.3 まとめ — 撮像は作り込める。判断は、継ぐしかない

外観検査AIの成否は、最新のAIモデルを買うことではなく、(1) 欠陥が確実に写る**撮像環境を作り込むこと**、(2) 学習データを丁寧に整え**AIの実力を育てること**、(3) 100%精度はない前提で**人とAIの分担を運用設計すること**という、技術以前の3点で決まる。AIは魔法の箱ではなく、これらの土台の上で初めて機能する道具である。

そして、この3点を整えるほど、残るものがはっきりする。**撮像は作り込める。データは整えられる。運用は設計できる。標準化できないものが、最後に一つ残る。**「これは見逃せない」という検査員の判断である。それは検査記録にも、しきい値にも、モデルの重みにも残らない。

撮像と運用に投資した企業は、外観検査AIを動かせる。だが本当に先行するのは、**その先に残った判断を引き出し、次の世代へ継げる企業**だ。

着手にあたっては、全工程を一斉にAI化しようとせず、最も困っている1工程・1製品で小さくPoCを始めることを推奨する。明日からの一歩は二つある。一つは、目視に最も負荷がかかっている検査工程を選び、その対象を「撮像できるか」から検証してみる。もう一つは、**AIが迷った境界事例を一つ選び、ベテランがなぜそう判定したのかを、理由ごと書き出してみる**ことである。前者がなければ、AIは動かない。後者がなければ、動いたAIは一代限りで終わる。

taiziii が提供する判断力の継承サービス SkillRelay（スキルリレー）では、検査基準書にもしきい値にもモデルの重みにも載らない「これは見逃せない」という熟練検査員の見極めを、なぜそう判断したのかという理由ごとAIで引き出し、帰属を保ったまま条件付きルールへ構造化し、組織の判断知として継承します。判断を平均したり、一本の正解に統合したりはしません。例外は例外として残します。AIは判断を代行せず、引き出し、分類し、問い返し、継承を助ける裏方に徹します。撮像・学習データ・運用設計といった土台づくりから、その先に残る判断の継承までを、ひと続きで扱える点に強みを持ちます。詳細：<https://taiziii.com/skillrelay/>