



Złożenie pracy online:
2025-09-23 10:55:45
Kod pracy:
59861/50106/CloudA

Marek Turcza
(nr albumu: 36177)

Praca magisterska

**„Organizacja i standaryzacja procesów etykietowania
danych wizyjnych w laboratorium - studium przypadku.”**

**„Organization and Standardization of Visual Data Annotation
Processes in a Laboratory - A Case Study.”**

Wydział: Wyższa Szkoła Biznesu - National-
Louis University

Kierunek: Zarządzanie

Specjalność: zarządzanie projektami

Promotor: dr hab. Wojciech Wdowiak

Pragnę serdecznie podziękować mojemu promotorowi, dr hab. Wojciechowi Wdowiakowi, za okazaną pomoc, cenne rady oraz merytoryczne wskazówki, które towarzyszyły mi na etapie przygotowywania i pisania niniejszej pracy magisterskiej. Jego wsparcie i zaangażowanie miały istotny wpływ na ostateczny kształt tej pracy.



Streszczenie

Celem pracy było zbadanie, jaki wpływ na funkcjonowanie jednostki laboratoryjnej ma usprawnienie procesów etykietowania danych. Badanie przeprowadzono w laboratorium działającym we współpracy firmy inżynierskiej z uczelnią techniczną. Zaproponowano zintegrowany łańcuch procesowy łączący BPM/BPMN, Lean/5S, QA/SPC (AQL, podwójne znakowanie, adjudykacja) oraz mechanizmy zgodności (RODO/GDPR, RBAC, ślad danych), działające w cyklu PDCA. Metodyka obejmowała mapowanie procesu, stabilizację wejścia (data readiness, IQA/SNR, kompletne metadane), limity WIP, selektywny pre-labeling i porównanie „przed-po” dla TAT, reworku i zgodności (κ/α). Wyniki: TAT 5,6→4,5 dnia (-19,6%), lead time P85 8,1→6,0 dnia (-26%), mediana cyklu 4,9→3,9 dnia (-20%), przepustowość +22%, rework 14,8%→9,1% (-5,7 p.p.), κ 0,73→0,87, α 0,78→0,90. Wnioski potwierdzają hipotezę główną: usprawnienia równocześnie skracają czas, ograniczają poprawki i podnoszą zgodność ocen, zapewniając audytowalność procesu. Przedstawiona Metodyka organizacji i standaryzacji procesów etykietowania stanowi praktyczne ramy wdrożeniowe.

Słowa kluczowe

Etykietowanie danych, Zarządzanie procesami, Zwinne metody zarządzania, Uczenie maszynowe, Optymalizacja, Studium przypadku, Laboratorium.



Abstract

The aim of this thesis was to examine the impact of process improvements in data annotation on the functioning of a laboratory unit. The study was conducted in a laboratory operating within a collaboration between an engineering company and a technical university. An integrated process governance framework was proposed, combining BPM/BPMN, Lean/5S, QA/SPC (AQL, double labeling, adjudication), and compliance mechanisms (GDPR, RBAC, data lineage), implemented in a PDCA cycle. The methodology included process mapping, input stabilization (data readiness, IQA/SNR, complete metadata), WIP limits, selective pre-labeling, and a before–after comparison for TAT, rework, and agreement (κ/α). Results: TAT 5.6→4.5 days (−19.6%), lead time P85 8.1→6.0 days (−26%), cycle time median 4.9→3.9 days (−20%), throughput +22%, rework 14.8%→9.1% (−5.7 p.p.), κ 0.73→0.87, α 0.78→0.90. The conclusions confirm the main hypothesis: process improvements simultaneously shorten execution time, reduce rework, and increase inter-rater agreement, while ensuring process auditability. The presented methodology of organizing and standardizing data annotation processes provides practical implementation guidelines.

Keywords

Data labeling, Process management, Agile management methods, Machine learning, Optimization, Case study, Laboratory.



Spis treści

Wykaz skrótów.....	3
Wstęp	5
Rozdział 1. Zarządzanie laboratorium etykietującego dane wizyjne	8
1.1. Struktura organizacyjna i procesy w laboratorium.....	8
1.2. Aktualnie wykorzystywane metody zarządzania procesami	16
1.3. Problemy i wyzwania z zarządzaniem laboratorium etykietującym dane wizyjne	21
1.4. Analiza rynku i konkurencyjności laboratorium	24
Rozdział 2. Istota i przebieg procesu etykietowania danych wizyjnych	30
2.1. Założenia i specyfika procesu oznaczania danych	31
2.2. Etykietowanie w ML/AI	33
2.2.1. Sztuczna inteligencja i uczenie maszynowe.....	34
2.2.2. Rola etykietowania danych	35
2.3. Wyzwania i oczekiwania wobec nowoczesnych modeli	37
2.4. Ochrona danych i polityka prywatności (RODO/PII, anonimizacja, kontrola dostępu).....	41
2.5. Definicja etykietowania/labelingu/annotation oraz etymologia.....	45
2.6. Historia rozwoju etykietowania (kamienie milowe, narzędzia, standaryzacja QA)	47
2.7. Czynniki wpływu i konieczność rozwoju	50
2.8. Zalety	51
2.9. Wady i ograniczenia / bariery.....	53
Rozdział 3. Propozycje i warianty optymalizacji procesów zarządzania laboratorium	56
3.1. Przegląd zwinnych metod zarządzania w kontekście etykietowania danych.....	56
3.2. Wybór metody zwinnego zarządzania odpowiedniej dla laboratorium	61
3.2.1. Kryteria oceny i ich operacjonalizacja	61
3.2.2. Ważenie kryteriów metodą AHP	62
3.2.3. Ocena wariantów i ranking bazowy	63
3.2.4. Rejestr ryzyk i mapa ryzyka.....	63
3.2.5. Korekta wyniku o ekspozycję na ryzyko	65
3.3. Optymalizacja przepływu pracy i redukcja kosztów.....	65
3.3.1. VSM to-be – cele przepływu i parametry strumienia.....	66
3.3.2. Czas taktu i balansowanie obciążenia	67
3.3.3. Automatyzacje i pre-labeling	67
3.3.4. Ekonomia usprawnień – koszt i zwrot	69
3.3.5. Harmonogram wdrożenia to-be (0–90 dni).....	69
3.4. Poprawa jakości etykietowania poprzez kontrolę i podwójną weryfikację	70
3.4.1. Kontrola próbkowa partii (AQL, ISO 2859-1).....	71
3.4.2. Podwójna weryfikacja i adjudkacja	71



3.4.3. Miary zgodności międzyoceniających i metryki dla obrazów	73
3.4.4. Jakość w środowisku zespołowym/crowd: złote zadania, adaptacja próby, konsensus	74
3.4.5. Kalibracje: rytm, artefakty i zamknięcie pętli	75
Rozdział 4. Wdrożenie wybranych metod zarządzania w laboratorium	77
4.1. Opis procesu wdrożenia	79
4.2. Zastosowane metody i narzędzia	84
4.3. Modelowanie procesu (BPMN as-is/to-be, punkty kontroli)	88
4.4. Analiza SWOT po wdrożeniu	91
4.5. Ocena wpływu optymalizacji na konkurencyjność i wydajność laboratorium	95
4.6. Metodyka organizacji i standaryzacji procesów etykietowania	103
Zakończenie	106
Bibliografia	110
Spis tabel	116
Spis rysunków	118
Spis wykresów	118



Wykaz skrótów

AQL - (ang. *Acceptable Quality Level*) - akceptowalny poziom jakości

BPM - (ang. *Business Process Management*) - zarządzanie procesami biznesowymi

BPMN - (ang. *Business Process Model and Notation*) – notacja modelowania procesów

COQ - (ang. *Cost of Quality*) – całkowity koszt jakości

ETL - (ang. *Extract–Transform–Load*) – pozyskiwanie i przetwarzanie danych

IQA - (ang. *Image Quality Assessment*) – ocena jakości obrazu

KPI - (ang. *Key Performance Indicator*) – kluczowy wskaźnik efektywności

MLOps - (ang. *Machine Learning Operations*) – operacjonalizacja uczenia maszynowego

PDCA - (ang. *Plan–Do–Check–Act*) – cykl doskonalenia

PII - (ang. *Personally Identifiable Information*) – dane osobowe

QA - (ang. *Quality Assurance*) – zapewnienie jakości

RACI - (ang. *Responsible–Accountable–Consulted–Informed*) – macierz odpowiedzialności

RBAC - (ang. *Role-Based Access Control*) – kontrola dostępu oparta na rolach

RODO/GDPR - (ang. *General Data Protection Regulation*) – rozporządzenie o ochronie danych osobowych

SLA - (ang. *Service Level Agreement*) – umowa o gwarantowanym poziomie usług

SNR - (ang. *Signal-to-Noise Ratio*) – stosunek sygnału do szumu

SOP - (ang. *Standard Operating Procedure*) – standardowa procedura operacyjna

SPC - (ang. *Statistical Process Control*) – statystyczna kontrola procesu

TAT - (ang. *Turnaround Time*) – czas realizacji

VSM - (ang. *Value Stream Mapping*) – mapowanie strumienia wartości

WIP - (ang. *Work In Progress*) – praca w toku

α - (ang. *alfa Krippendorffa*) – miara zgodności międzyoceniających dla wielu kategorii/skal

κ - (ang. *kappa Cohena*) – miara zgodności międzyoceniających

5S - (ang. *Sort, Set in order, Shine, Standardize, Sustain*) – metoda 5S: selekcja, systematyka, sprząatanie, standaryzacja, samodyscyplina



AI - (ang. *Artificial Intelligence*) – sztuczna inteligencja

API - (ang. *Application Programming Interface*) – interfejs programistyczny aplikacji

CV - (ang. *Computer Vision*) – widzenie komputerowe

DCM - (ang. *Document Control Management*) – zarządzanie kontrolą dokumentów

DQ - (ang. *Data Quality*) – jakość danych

ML - (ang. *Machine Learning*) – uczenie maszynowe

TCO - (ang. *Total Cost of Ownership*) – całkowity koszt posiadania

Labeling - (ang. *data labeling/annotation*) – etykietowanie (adnotowanie) danych

pre-labeling – wstępne etykietowanie (wstępne adnotowanie)

rework - prace korygujące / ponowne wykonanie (poprawki)



Wstęp

Etykietowanie danych wizyjnych jest kluczowym ogniwem łańcucha wartości sztucznej inteligencji. To właśnie dyscyplina pracy, możliwość audytu i jednoznaczność wytycznych przekładają się na wiarygodność rezultatów, a zwinne zarządzanie pozwala skrócić czas realizacji bez pogorszenia jakości. Temat łączy perspektywę ekonomiczną (produktywność, całkowity koszt jakości, TAT, ryzyko niezgodności) z praktykami operacyjnymi, pokazując, że o skuteczności modeli decyduje nie tylko algorytm, lecz także jakość i organizacja danych.

Praca nosi tytuł: „Organizacja i standaryzacja procesów etykietowania danych wizyjnych w laboratorium – studium przypadku”.

Obszarem badań jest laboratorium etykietujące dane wizyjne, a przedmiotem badań – procesy etykietowania i towarzyszący im ład zarządczy. Kontekst empiryczny stanowi laboratorium prowadzone we współpracy firmy inżynierskiej z uczelnią techniczną, co zapewnia osadzenie rozważań w realnych uwarunkowaniach operacyjnych.

Celem pracy jest zaprojektowanie i uzasadnienie spójnego ładu procesowego dla etykietowania danych wizyjnych oraz weryfikacja, czy jego wdrożenie w środowisku laboratoryjnym skraca czas realizacji (TAT) i ogranicza rework przy jednoczesnym utrzymaniu lub wzroście zgodności międzyoceniających (κ , α). Cel teoretyczny obejmuje integrację podejść BPM, Lean i QA z cyklem PDCA oraz dobór miar zgodności; cel praktyczny – zaprojektowanie i wdrożenie usprawnień wraz z zasadami dokumentowania i wersjonowania wytycznych.

Główne pytanie badawcze brzmi: Jaki wpływ na funkcjonowanie jednostki laboratorium ma usprawnienie procesów etykietowania danych?

W pracy operacjonalizuję „funkcjonowanie” poprzez cztery wymiary oceny: sprawność operacyjną (TAT, przepustowość), jakość (κ/α , rework), zgodność i bezpieczeństwo (RODO/GDPR, ślad danych) oraz efektywność ekonomiczną (COQ/TCO).

Pytania szczegółowe obejmują:

1. Jakie są przejawy braku spójności architektury procesu i dokumentacji oraz ich wpływ na rework i zgodność (RACI, SOP, KPI; BPMN, DCM, Git)?
2. Jak stabilizować jakość wejścia przy zmienności akwizycji obrazu (SOP, IQA, SNR; kompletność metadanych i ETL/DQ)?



3. Które mechanizmy najsilniej skracają czas cyklu etykietowania (WIP, standaryzacja pracy, VSM, 5S; półautomatyzacje, MLOps, API)?
4. Jak zaprojektować kontrolę zgodności i bezpieczeństwa danych (RODO/GDPR, PII, RBAC; ślad danych i audyt)?
5. Jak zorganizować kontrolę jakości i kalibracje, aby utrzymać wysoką zgodność międzyoceniających (AQL, QA, PDCA; κ , α)?

Z pytaniami powiązano hipotezy:

Hipoteza główna:

Usprawnienie procesów etykietowania danych w jednostce laboratoryjnej - poprzez zintegrowany łańd procesowy oraz standaryzację akwizycji i etykietowania - istotnie poprawia funkcjonowanie laboratorium: skraca czas realizacji (TAT) i zwiększa przepustowość, redukuje rework, podnosi zgodność międzyoceniających (κ/α), wzmacnia zgodność regulacyjną i audytowalność procesu oraz obniża koszt jakości (COQ/TCO) względem stanu wyjściowego.

Obok hipotezy głównej sformułowano zestaw hipotez uzupełniających (H1–H5), odpowiadających kolejno pierwszemu, drugiemu, trzeciemu, czwartemu i piątemu pytaniu badawczemu. Każda hipoteza dotyczy odrębnego wymiaru funkcjonowania laboratorium - architektury i dokumentacji procesu, jakości wejścia/akwizycji, skrócenia czasu cyklu, zgodności i bezpieczeństwa danych oraz zapewnienia jakości i kalibracji - i jest weryfikowana na podstawie porównań „przed-po” z użyciem miar TAT, udział reworku, κ/α , przepustowość oraz wskaźników zgodności/audytu.

Brzmienie hipotez jest następujące: H1–H5.

H1 – Uporządkowanie map procesów oraz nadzór zmian i aktualizacja wytycznych (DCM) obniża rework i liczbę niezgodności;

H2 – standaryzacja akwizycji obrazu oraz kryteria odrzutu materiałów suboptymalnych zwiększają κ i zmniejszają TAT;

H3 – ograniczenie WIP i standaryzacja pracy skracają TAT co najmniej o 15% bez pogorszenia jakości;

H4 – wdrożenie mechanizmów zgodności (RODO/PII, RBAC, ślad danych) skraca czas obsługi zgłoszeń i redukuje ryzyko niezgodności;

H5 – podwójna weryfikacja i cykliczne kalibracje zwiększają κ i stabilizują wyniki QA.

Metodyka postępowania badawczego obejmuje analizę dokumentów i dzienników procesowych, mapowanie BPMN, VSM i pomiar TAT, wykorzystanie



wykresów SPC, plan kontroli próbkowej (AQL/ISO 2859-1), podwójne znakowanie i adjudykację oraz estymację κ Cohena i α Krippendorffa; ocenę efektów przeprowadza się w układzie porównań przed–po wdrożeniu dla TAT, reworku i κ . W kolejnych częściach przedstawiono uwarunkowania organizacyjne laboratorium, istotę i przebieg procesu etykietowania w kontekście ML/AI, propozycję usprawnień wraz z uzasadnieniem oraz wyniki wdrożenia w formule studium przypadku.



Rozdział 1. Zarządzanie laboratorium etykietującego dane wizyjne

Laboratorium funkcjonuje w otoczeniu rosnącego popytu na dane wysokiej jakości dla modeli ML, przy presji kosztowej i krótkich cyklach wdrożeń. W tym kontekście o przewadze decydują dojrzałe struktury organizacyjne i zarządcze, które stabilizują jakość, skracają czas realizacji oraz zwiększają przejrzystość odpowiedzialności (Mikuła, Gawron, 2020).

1.1. Struktura organizacyjna i procesy w laboratorium

Laboratorium etykietujące dane wizyjne należy postrzegać jako organizację usługową, której cele i strategie przejawiają się poprzez fundamentalne funkcje zarządzania: planowanie, organizowanie, kierowanie i kontrolowanie przepływów pracy (Mikuła, Gawron, 2020). Perspektywa ta podkreśla, że zasady projektowania organizacji - takie jak podział pracy, departamentalizacja, hierarchia uprawnień i kanały komunikacji - odgrywają kluczową rolę w kształtowaniu skuteczności i efektywności realizacji zleceń adnotacyjnych (Gembala, 2013). Zasady te są namacalne; są ucieleśnione w konkretnych uprawnieniach, obowiązkach i interakcjach między rolami, które znacząco wpływają na jakość i terminowość dostaw. Przyjęcie podejścia zorientowanego na procesy wymaga, aby każda funkcja operacyjna była sformułowana jako seria działań charakteryzujących się nakładami, nakładami oraz wyznaczonym właścicielem procesu. W kontekście laboratorium usługowego o dużej intensywności wiedzy, pociąga to za sobą integrację personelu, procedur i narzędzi informatycznych w ramach jednolitego systemu zarządzania. System ten jest cyklicznie wzmacniany poprzez mechanizmy kontroli i uczenia się organizacyjnego (Mikuła, Gawron, 2020).

W laboratorium akredytowanym lub dążącym do akredytacji struktura ma dodatkowy wymiar: wspiera bezstronność i rzetelność wyników. Badania ankietowe w polskich laboratoriach wskazują, że aż 75% jednostek uznaje strukturę za czynnik ułatwiający zarządzanie i zapewniający bezstronność, a ponad połowa dostosowywała układ organizacyjny do wymagań norm (Mytych, Ligarski, 2017). Adaptacje obejmowały tworzenie dodatkowych komórek, doprecyzowanie procedur i zdefiniowanie ról zgodnych z wymogami systemów jakości. W praktyce warto rozdzielać nadzór nad jakością od produkcji usługowej, by ograniczyć konflikt interesów. Takie rozróżnienie roli „właściciela procesu” i roli „quality assurance” wzmacnia przejrzystość



i odpowiedzialność. Wnioski te są spójne z podejściem systemowym do zarządzania jakością w laboratoriach (Mytych, Ligarski, 2017).

Ramy zarządzania procesami etykietowania w całym cyklu - obejmującym planowanie standardów, wdrażanie, weryfikację i działania korygujące - zapewniają kompleksowe zarządzanie jakością (TQM) oraz podejście PDCA (Skubała, 2023). W warunkach laboratoryjnych ramy te są stosowane do planowania danych i wymagań dotyczących etykietowania, wykonywania adnotacji, sprawdzania dokładności i zgodności oraz rewizji instrukcji i szkoleń. Włączając PDCA do codziennych praktyk, takich jak przeglądy, audyty wewnętrzne i retrospektywy, organizacje mogą stabilizować wyniki i systematycznie zmniejszać zmienność. To zwiększenie przewidywalności jakości jest niezbędne dla rozległych łańcuchów uczenia się modeli. Ponadto cykl PDCA ułatwia wdrażanie działań zapobiegawczych przed pojawieniem się niezgodności, stanowiąc podstawę doskonalenia operacyjnego w usługach opartych na wiedzy (Skubała, 2023).

Procedura etykietowania, stanowiąca podstawę operacji, powinna być ustrukturyzowana jako sekwencja obejmująca dobrze zdefiniowane dane wejściowe (dane, specyfikacje, instrukcje), dane wyjściowe (etykiety, wskaźniki jakości), zasoby (narzędzia) i punkty kontrolne (Mikuła, Gawron, 2020). Niezbędne jest wyznaczenie właścicieli procesów, którzy posiadają uprawnienia decyzyjne i odpowiadają za kluczowe wskaźniki efektywności (KPI). Organizując interfejsy między procesami, można zminimalizować straty podczas przekazywania zadań. Opisy procesów powinny być wystarczająco szczegółowe, aby ułatwić szkolenie i zapewnić powtarzalność, a jednocześnie wystarczająco elastyczne, aby uwzględniać zmieniające się zamówienia. Takie podejście umożliwia transfer wiedzy z poziomu eksperckiego na poziom organizacyjny, zwiększając tym samym zdolność do skalowania operacji (Mikuła, Gawron, 2020).

W obszarze praktyki procesowej utrzymanie spójności dokumentacji ma ogromne znaczenie, obejmując wszystko, od polityk i procedur po instrukcje robocze i zapisy kontroli jakości (Skubała, 2023). Dokumentacja powinna usprawniać działania, a nie je zastępować; służy ona jako środek przekazywania standardów i dostarczania dowodów ich przestrzegania. Skutecznie ustrukturyzowane ramy dokumentacji zmniejszają prawdopodobieństwo rozbieżności w interpretacji i wspomagają szkolenia. W środowiskach, w których wytyczne dotyczące adnotacji często ewoluują, wskazane jest wdrożenie krótkich cykli aktualizacji wraz z przeglądami kierownictwa. Audyty



wewnętrzne i oceny kierownictwa uzupełniają proces uczenia się. Ta koncepcja „jakości poprzez dokumentację” sprzyja kulturze odpowiedzialności (Skubała, 2023).

Warstwa metryk (KPI), powinna obejmować zarówno jakość, jak i wydajność, w tym takie czynniki, jak przepustowość, czas cyklu, tempo poprawek, spójność między annotatorami oraz czas reakcji na zmiany w specyfikacjach (Mikuła i Gawron, 2020). Aby skutecznie mierzyć wydajność, niezbędne jest ustalenie źródeł danych i przypisanie odpowiedzialności za raportowanie. Metryki powinny być dostosowane od poziomu strategicznego do poziomu zespołu, zapewniając, że każda osoba rozumie, jak jej wkład wpływa na ogólne wyniki. Wiarygodne KPI stanowią podstawę przeglądów Plan-Do-Check-Act (PDCA) i wpływają na decyzje dotyczące zmian w organizacji. Nadmiar metryk bez realnych rezultatów może prowadzić do demotywacji; dlatego należy skupić się na „kilku, ale kluczowych” wskaźnikach. Ta dyscyplina dotyczy informacji zarządczej, a nie statystyki dla samej statystyki (Mikuła i Gawron, 2020).

Wybrany poziom centralizacji zależy zarówno od złożoności portfela zamówień, jak i doświadczenia zespołu; centralizacja sprzyja jednolitości, a decentralizacja umożliwia szybkie reagowanie (Gembala, 2013). W kontekście laboratorium etykietowania często zaleca się utworzenie „centrum kompetencji jakościowych”, ponieważ konsoliduje ono standardy i audyty, umożliwiając jednocześnie zespołom procesowym zarządzanie planowaniem operacyjnym. Takie podejście skutecznie łączy spójność metodologiczną z lokalną reakcją. Niezbędne jest określenie jasnych granic dla procedur podejmowania decyzji i eskalacji, ponieważ wszelkie niejasności w tych obszarach mogą prowadzić do opóźnień i „pustych przebiegów” informacji. Struktura organizacyjna powinna wspierać przepływ wartości, a nie odwrotnie (Gembala, 2013).

Badania w laboratoriach wskazują, że dostosowanie struktury organizacyjnej do wymogów norm i procesów zwiększa zarówno wydajność, jak i obiektywizm (Mytych, Ligarski, 2017). Zazwyczaj te modyfikacje pociągają za sobą ustanowienie ról i jednostek, które wspierają jakość, doprecyzowanie obowiązków oraz formalizację przepływu informacji. W rezultacie zmniejsza się prawdopodobieństwo błędów systemowych, a jednocześnie poprawia się przejrzystość procesu decyzyjnego. Zatem struktura służy jako środek zapewniający kompetencje organizacyjne, a nie jako cel ostateczny. Zasada ta ma zastosowanie zarówno w laboratoriach badawczych i diagnostycznych, jak i w usługach przetwarzania danych. Wskazane jest przełożenie tych koncepcji na praktyczne mapy ról i interfejsów (Mytych, Ligarski, 2017).



Kluczowym elementem systemu jest dobrze rozwinięty proces zarządzania skargami i zażaleniami klientów, który przyczynia się do cyklu doskonalenia (Wierzowiecka, 2017). Norma ISO 10002 oferuje wytyczne dotyczące tworzenia takiego procesu, obejmującego etapy od akceptacji i oceny, przez udzielenie odpowiedzi, zamknięcie sprawy, po analizę przyczyn źródłowych. Warto zauważyć, że norma ISO/IEC 17025, choć nakazuje efektywną współpracę z klientami, nie określa konkretnych rozwiązań operacyjnych; w tym zakresie ISO 10002 stanowi niezbędne uzupełnienie. Dla laboratorium usług danych każda skarga stanowi istotne źródło informacji na temat rozbieżności w standardach, szkoleniach i narzędziach. Monitorowanie czasu reakcji, satysfakcji klienta i skuteczności działań naprawczych buduje zaufanie. W rezultacie skargi przekształcają się w cenne dane jakościowe, a nie są jedynie postrzegane jako „koszt” (Wierzowiecka, 2017).

Skuteczne rozpatrywanie skarg wymaga wsparcia strukturalnego, które obejmuje zaangażowanie kadry kierowniczej wyższego szczebla, szkolenie personelu oraz przeprowadzanie audytów i przeglądów procesów (Wierzowiecka, 2017). W praktyce kluczowe jest określenie ról: obowiązki na pierwszej linii obejmują przyjmowanie i klasyfikację, role analityczne koncentrują się na badaniu przyczyn źródłowych, podejmowanie decyzji dotyczy zatwierdzania działań, a komunikacja obejmuje informację zwrotną. Ciągłe monitorowanie wskaźników związanych z procesem rozpatrywania skarg ułatwia ocenę jego skuteczności i identyfikację obszarów wymagających poprawy. Ponadto służy jako mechanizm wczesnego ostrzegania o każdym spadku standardu adnotacji. Takie podejście gwarantuje, że „głos klienta” jest konsekwentnie uwzględniany w cyklu PDCA laboratorium, co zapewnia jakościową przewagę w zakresie usług danych (Wierzowiecka, 2017).

Podejście procesowe wykracza poza samo tworzenie mapy; obejmuje również ewolucję organizacji w kierunku zarządzania procesami biznesowymi (BPM) i zwiększania jej dojrzałości (Winiowski, 2016). W literaturze wyróżnia się różne stopnie „myślenia procesowego”, od funkcjonalnych „silosów” po organizację w pełni zorientowaną na procesy. Każdy stopień niesie ze sobą określone konsekwencje dla struktury, kompetencji i wykorzystywanych narzędzi. W laboratorium usługowym, rozwój organizacji wzdłuż tego kontinuum sprzyja większej przejrzystości w zakresie odpowiedzialności i skraca czas trwania cykli decyzyjnych. Niemniej jednak, osiągnięcie tego wymaga jednoczesnej transformacji zarówno kultury organizacyjnej, jak i systemów



informatycznych. Podejście systemowe okazuje się skuteczniejsze niż selektywne wdrożenia (Winiowski, 2016).

Współczesne podejście do zarządzania procesami, często określane jako „trzecia fala” BPM, łączy postęp technologiczny z dyscypliną procesową, koncentrując się na modelowaniu, monitorowaniu i optymalizacji w pętli zamkniętej (Jasińska, 2017). W kontekście laboratorium etykietowania przekłada się to na wykorzystanie cyfrowych modeli procesów, dzienników zdarzeń, pulpity KPI oraz eksperymentów ukierunkowanych na szybkie doskonalenie. Zwinne techniki zarządzania idealnie wpisują się w te ramy, sprzyjając połączeniu iteracji i przejrzystości. Nacisk kładziony jest na integrację osób, aplikacji i danych w strumieniu wartości. W rezultacie struktura organizacyjna służy jako czynnik ułatwiający procesy, a nie jako ich przeszkoda. Takie podejście toruje drogę do skalowalności i stabilności jakości (Jasińska, 2017).

Zarządzanie Lean Management w sektorach usługowych, w tym w laboratoriach, oferuje skuteczne strategie minimalizacji marnotrawstwa, skrócenia czasu oczekiwania i zapewnienia stabilnego przepływu pracy (Bąk, 2022). Zastosowanie narzędzi takich jak standaryzacja procesów pracy, techniki zarządzania wizualnego, wyrównywanie obciążenia i mapowanie strumienia wartości poprawia jakość i skraca czas cyklu. Studia przypadków w sektorze opieki zdrowotnej pokazują, że wdrożenie praktyk Lean przynosi znaczące usprawnienia operacyjne i jakościowe, nawet w kontekście zadań specjalistycznych. W laboratorium etykietującym, zastosowanie standardowych etykiet, list kontrolnych, wizualnych tablic statusu i spójnego rytmu kontroli jakości przynosi porównywalne rezultaty. Kluczem do sukcesu jest dostosowanie tych narzędzi do konkretnego środowiska, a nie jedynie ich powielanie. Lean reprezentuje podejście skoncentrowane na przepływie, a nie jedynie na zbiorze procedur (Bąk, 2022).

Związek między strukturą a procesem jest wzajemny: strumień wartości powinny być odzwierciedlone w strukturze, podczas gdy procesy muszą wykorzystywać dobrze zdefiniowane ścieżki decyzyjne i informacyjne (Gembala, b.d.). Projektując rolę, kluczowe jest uwzględnienie powiązań w zakresie raportowania, informacji i wsparcia, aby zapewnić podejmowanie decyzji jak najbliżej poziomu operacyjnego. Organizacje o płaskiej strukturze wymagają wyższych standardów i samoregulacji, podczas gdy te o rozbudowanej strukturze wymagają bardziej sformalizowanych procedur i zdefiniowanych ról pracowniczych. W obu scenariuszach korzystne jest minimalizowanie „tarcia organizacyjnego” na styku ról. Skuteczne praktyki obejmują



jasne definicje ról, macierze RACI i mapy interfejsów, które ostatecznie prowadzą do zmniejszenia zakłóceń w przepływie pracy (Gembala, b.d.).

Funkcje zarządzania muszą być spójne w różnych obszarach: wyznaczanie celów jakościowych poprzez planowanie, organizacja zasobów, kierowanie zespołami oraz kontrola rezultatów (Mikuła, Gawron, 2020). W kontekście usług adnotacji, planowanie obejmuje definiowanie standardów etykietowania i kryteriów akceptacji. Organizacja wiąże się z tworzeniem ram decyzyjnych i harmonogramowaniem dostaw. Przywództwo ukierunkowane jest na rozwijanie kompetencji i wspieranie motywacji w zadaniach, które mogą być monotonne i powtarzalne, ale mają istotną wartość biznesową. Kontrola - obejmująca nadzór - dostarcza niezbędnych informacji zarządczych do efektywnego zarządzania portfelem zamówień i obciążeniem pracą. Ta „czwórka” działa w sposób cykliczny, a nie liniowy (Mikuła, Gawron, 2020).

W laboratorium architektura jakości wykracza poza zwykłe inspekcje końcowe; obejmuje kompleksowy system obejmujący planowanie, kontrolę, zapewnienie jakości i doskonalenie (Skubała, 2023). W praktyce oznacza to ustanowienie jasno zdefiniowanych bramek jakości (QA gates), a także wdrożenie procedur pobierania próbek i podwójnej kontroli w obszarach o podwyższonym ryzyku błędu. Niezbędna jest ciągła kalibracja, składająca się z krótkich sesji mających na celu ujednoczenie interpretacji standardów wśród annotatorów. Istotne jest, aby dane dotyczące jakości były przekazywane z powrotem właścicielom procesów i autorom instrukcji. Ponadto, ciągłe przeglądy jakości z udziałem kierownictwa są niezbędne, aby zapewnić skuteczne utrzymanie priorytetów i zasobów. Tylko dzięki takiemu podejściu jakość można postrzegać jako nieodłączną cechę systemu, a nie jako produkt indywidualnych heroicznych działań (Skubała, 2023).

Skargi i zażalenia stanowią istotny kanał edukacyjny dla organizacji, dostarczając cennych informacji na temat jakości i spójności procesów (Wierzowiecka, 2017). Wytyczne ISO 10002 podkreślają znaczenie przejrzystości, uczciwości proceduralnej i czasu reakcji – czynników równie istotnych w przypadku usług danych. Formalizacja procesu rozpatrywania skarg pozwala organizacjom na przeprowadzanie analiz przyczyn źródłowych i skuteczne ustalanie priorytetów działań naprawczych. Efektywne zarządzanie skargami wzmacnia zaufanie klientów i zwiększa wskaźniki retencji. W kontekście zarządzania portfelem, skargi ułatwiają segmentację ryzyka. Ten aspekt stanowi dodatkowy wymiar „głosu klienta” w ramach metody PDCA (Wierzowiecka, 2017).



Aby stać się organizacją zorientowaną na procesy, organizacja musi przejść jednoczesną transformację kulturową i technologiczną: obejmuje to modelowanie procesów, monitorowanie w czasie rzeczywistym oraz integrację aplikacji (Jasińska, 2017). W obszarze zarządzania procesami biznesowymi (BPM) proces jest postrzegany jako „produkt”, który podlega projektowaniu, wdrażaniu i ciągłej optymalizacji. W laboratorium etykietowania, podstawowy przepływ zgłoszeń można usprawnić poprzez włączenie dzienników zdarzeń, które ułatwiają odkrywanie procesów, powszechnie znane jako eksploracja procesów. Takie usprawnienia gwarantują, że decyzje dotyczące zmian są podejmowane na podstawie danych, a nie jedynie intuicji. Z czasem organizacja ewoluuje w kierunku dojrzałości, przechodząc od map „na ścianie” do stosowania pętli sterowania opartych na danych. Ta ewolucja stanowi fundament skalowania zwinnego (Jasińska, 2017).

Wdrażanie metod Lean/BPM często napotyka na typowe przeszkody, takie jak opór przed zmianami, problemy z komunikacją, przeciążenie oraz trudności w dopasowaniu struktur do przepływów pracy (Bąk, 2022). W dostępnej literaturze podkreśla się znaczenie przywództwa klinicznego i technicznego, szczególnie w rolach liderów domeny, a także konieczność stosowania rotacyjnych systemów pracy i praktykę „Gemba Walk” w celu identyfikacji marnotrawstwa. W obszarze usług danych analogiczne praktyki obejmują inspekcje stanowisk pracy, oceny zaległości oraz obserwacyjne badania procesów pracy „time & motion” przeprowadzane bez stygmatyzacji. Budowanie poparcia dla zmian ułatwiają transparentne rezultaty, takie jak skrócenie cykli i mniejsza liczba niekompletnych wpisów. Aby uchronić się przed „teatrem Lean”, niezbędna jest cierpliwość i iteracyjny przepływ pracy. Osiągnięcie trwałych rezultatów wymaga dostosowania narzędzi do nadrzędnych celów (Bąk, 2022).

Kompetencje mają równie istotne znaczenie; profil roli obejmuje nie tylko adnotatorów, ale także trenerów standardów, audytorów jakości, właścicieli procesów i analityków danych (Mikuła, Gawron, 2020). Aby zminimalizować rozbieżności interpretacyjne i zwiększyć spójność etykiet, niezbędne są systematyczne szkolenia i kalibracje. Co więcej, zadania wymagające dużej wiedzy wymagają skupienia się na motywacji - strukturyzacji pracy w taki sposób, aby łączyć powtarzalność z możliwościami rozwoju. Kluczowe jest, aby kontrola była postrzegana jako forma wsparcia, a nie jedynie mechanizm wykrywania błędów. Nieprzyjęcie tej perspektywy może prowadzić do wzrostu ukrytej rotacji i spadku jakości. „Miękkie” elementy systemu odgrywają kluczową rolę w określaniu „twardych” rezultatów (Mikuła i Gawron, 2020).



W kontekstach charakteryzujących się znaczną niepewnością procesową, korzystne okazuje się wdrożenie iteracyjnej organizacji pracy: krótkich sprintów, ewaluacji i retrospektyw, obejmujących zarówno jakość, jak i operacje (Jasińska, 2017). Takie podejście ustanawia rytm synchronizacji, który obejmuje aktualizację instrukcji, rekaliibrację próbek i podejmowanie decyzji dotyczących modyfikacji narzędzi. Uwzględnienie informacji zwrotnej przyspiesza identyfikację i korygowanie rozbieżności. Zespoły interdyscyplinarne, w których jakość jest zintegrowana w ramach zespołu, a nie tylko funkcjonuje „obok”, zmniejszają tendencję do projektowania problemów. Na poziomie portfela strategia ta poprawia przewidywalność realizacji. Zwinność służy jako środek do osiągnięcia stabilności, a nie jej przeciwdziałania (Jasińska, 2017).

Do projektowania relacji z klientem należy podchodzić w sposób zorientowany na proces, uwzględniając przejrzyste umowy o gwarantowanym poziomie usług (SLA), jasne definicje „gotowości wejściowej”, ustrukturyzowany format raportowania zmian oraz efektywny system zarządzania reklamacjami (Wierzowiecka, 2017). Ustalenie jasnych kryteriów wejściowych, takich jak bramki jakościowe dla danych źródłowych, minimalizuje występowanie niekompletnych zamówień i konieczność wprowadzania poprawek. Ponadto formalne ścieżki zmian zakresu chronią przepływ pracy przed „odchyleniem od wymagań”. Wspólne przeglądy metryk i reklamacji zapewniają, że inicjatywy usprawniające koncentrują się na obszarach, w których przyniosą największe korzyści. Takie podejście wzmacnia również reputację laboratorium jako wiarygodnego i elastycznego partnera. Takie zarządzanie relacjami jest integralnym elementem systemu jakości (Wierzowiecka, 2017).

Struktura organizacyjna laboratorium etykietującego dane wizyjne musi być ściśle powiązana z zasadami logiki procesowej i zarządzania jakością. To powiązanie powinno opierać się na cyklu PDCA, kompleksowej dokumentacji i odpowiednich metrykach, a także być uzupełnione o proces reklamacyjny zgodny z normą ISO 10002 (Skubała, 2023; Wierzowiecka, 2017). Badania wskazują, że celowe dostosowanie tej struktury znacząco poprawia zarówno bezstronność, jak i skuteczność (Mytych, Ligarski, 2017). Trzecia fala metodologii Business Process Management (BPM) i Lean oferuje narzędzia ułatwiające iteracyjną optymalizację przepływów pracy i redukcję zmienności (Jasińska, 2017; Bąk, 2022). Skuteczność tych narzędzi zależy od kompetencji i kultury organizacyjnej; bez tych elementów struktura jest jedynie koncepcją teoretyczną. Całe ramy powinny charakteryzować się podejściami opartymi na danych i uczeniem się



w pętli zamkniętej. System skonstruowany w ten sposób zwiększa konkurencyjność, przewidywalność i ogólną jakość wyników etykietowania.

1.2. Aktualnie wykorzystywane metody zarządzania procesami

Podejście procesowe „trzeciej fali” BPM integruje modelowanie, monitorowanie i iteracyjne udoskonalenia oparte na danych; w kontekście laboratorium etykietowania skutkuje to krótkimi cyklami sprzężenia zwrotnego, które zwiększają jakość i terminowość partii danych (Jasińska, 2017). Faza początkowa obejmuje organizację architektury procesów, która obejmuje identyfikację procesów kluczowych (adnotacje), działania zarządcze (planowanie, przeglądy) oraz funkcje wsparcia (szkolenia, IT, zgodność). Ten etap obejmuje przypisanie odpowiedzialności, a także zdefiniowanie danych wejściowych, wyjściowych i kluczowych wskaźników efektywności (KPI). Transformacja przebiega w kilku etapach: począwszy od identyfikacji i mapowania, poprzez wdrożenie i pomiar, aż po ciągłą optymalizację - zapewniającą kaskadowość i mierzalność celów (Winiowski, 2016).

Cykl PDCA działa jak „metronom jakości”, obejmujący planowanie standardów i partii danych (P), przeprowadzanie adnotacji i inspekcji (D), weryfikację wskaźników i przeprowadzanie audytów (C), a następnie inicjowanie działań korygujących i doskonalących (A). Takie podejście przekształca przeglądy zarządzania w mechanizm uczenia się, a nie jedynie w kontrolę zgodności. Co istotne, model PDCA jest zgodny z podejściem zwinnym, zapewniając stały rytm sprintów i retrospektyw (Skubała, 2023).

Ramy normatywne ustanowione w normie ISO/IEC 17025:2017, zgodne z normą ISO 9001:2015, wzmacniają te ujęcie, nadając priorytet podejściu zorientowanemu na proces, wnioskowaniu opartemu na ryzyku i mierzalnym rezultatom. W praktyce oznacza to wdrożenie rejestru ryzyka (na przykład dryfu wytycznych i przeciążenia QA), wykorzystanie macierzy reakcji oraz walidację modyfikacji instrukcji, co przyczynia się do stabilizacji jakości w warunkach wahań w portfelu zamówień (Gruszka, 2018).

Jednocześnie przepływ pracy optymalizuje się narzędziami Lean: 5S, Kanban, VSM, Takt i Gemba. Wdrożenia usługowe (m.in. w ochronie zdrowia) pokazują spadek czasu cyklu, mniejszą zmienność i lepszą widoczność wąskich gardeł; w „etykietowaniu jako usłudze” przekłada się to na limity WIP, przejrzyste kolejki i standaryzację pracy (Bąk, 2022).

Aby zapewnić spójność w zarządzaniu, niezbędne jest dostosowanie metod do różnych poziomów zarządzania: strategii, jakości, doskonalenia, zgodności, relacji



z klientami i zarządzania wiedzą. Jak ilustruje poniższa tabela 1, istnieje związek między „warstwa → metoda → cel → artefakt → KPI”, który wspomaga kaskadowanie celów i eliminację nieskutecznej dokumentacji (Winiowski, 2016).

Tabela 1. Mapowanie metod na cele i KPI

Warstwa ładu	Metoda/ramy	Cel zarządczy	Artefakt/Praktyka	Przykładowy KPI
Strategia i cele	BPM (trzecia fala), mapa procesów	Powiązanie celów z przepływami pracy	Mapa procesów, słownik wejść/wyjść	Terminowość (SLA)
Zarządzanie jakością	PDCA, audyty wewnętrzne	Stabilizacja jakości, wczesne wykrywanie odchyleń	Plan jakości, przeglądy kierownictwa	Odsetek reworku [%]
Doskonalenie operacyjne	Lean (5S, VSM, Kanban)	Skrócenie czasu cyklu, eliminacja strat	Tablice Kanban, standard pracy	Czas cyklu [dni]
Bezpieczeństwo i zgodność	ISO/IEC 17025 + ISO 9001	Bezstronność, myślenie o ryzyku	Rejestr ryzyk, działania korygujące	Niezgodności/audyt [#]
Relacja z klientem	ISO 10002 (reklamacje)	Głos klienta w PDCA	Procedura skarg, metryki satysfakcji	Czas obsługi skargi [dni]
Zarządzanie wiedzą	Repozytoria, wersjonowanie wytycznych	Spójność interpretacji etykiet	Księga adnotacji, rejestr zmian	Zgodność κ

Źródło: opracowanie własne na podstawie: Jasińska (2017); Winiowski (2016); Gruszka (2018); Wierzowiecka (2017)

Aby zapewnić jednolitość interpretacji, niezbędne jest posiadanie „żywej” dokumentacji, która obejmuje księgę etykiet, listy kontrolne zapewnienia jakości, rejestry zmian i zasady wersjonowania. Dokumenty te powinny być regularnie aktualizowane podczas przeglądów i retrospektyw oraz niezwłocznie udostępniane zespołom (Skubała, 2023).



Warstwa QA musi integrować ewaluację próbek, podwójne tagowanie dla kategorii wysokiego ryzyka oraz rozwiązywanie sporów, a także proste kryteria walidacji (takie jak zachowanie spójności formatu). Informacje pochodzące z QA i reklamacji powinny przyczyniać się do cyklu PDCA, a podstawowe wykresy kontrolne powinny ułatwiać wczesną identyfikację odchyień od jakości. Takie podejście gwarantuje, że jakość jest nieodłącznie włączona w proces, a nie jedynie dodawana na końcu (Skubała, 2023; Wierzowiecka, 2017).

Kluczowe Wskaźniki Efektywności (KPI) powinny być rozpowszechniane kaskadowo: począwszy od umów o gwarantowanym poziomie usług (SLA) i satysfakcji klienta na poziomie strategicznym, poprzez cykle i poprawki na poziomie procesów, aż po produktywność i spójność między annotatorami κ na poziomie zespołu. Skuteczność tych wskaźników zależy od ich jasnej interpretacji i widoczności w komisjach ds. przeglądu, co umożliwi szybkie i proporcjonalne działania naprawcze (Winiowski, 2016).

Biorąc pod uwagę, że jakość etykiet zależy od jakości danych wejściowych, kluczowe jest, aby standardy akwizycji obrazu - obejmujące oświetlenie, optykę i metadane - a także kryteria odrzucenia materiałów suboptymalnych stanowiły integralne elementy systemu zarządzania. Elementy te przyczyniają się do stabilizacji parametrów procesu i zwiększają interpretowalność kluczowych wskaźników efektywności (KPI) (Jasińska, 2017).

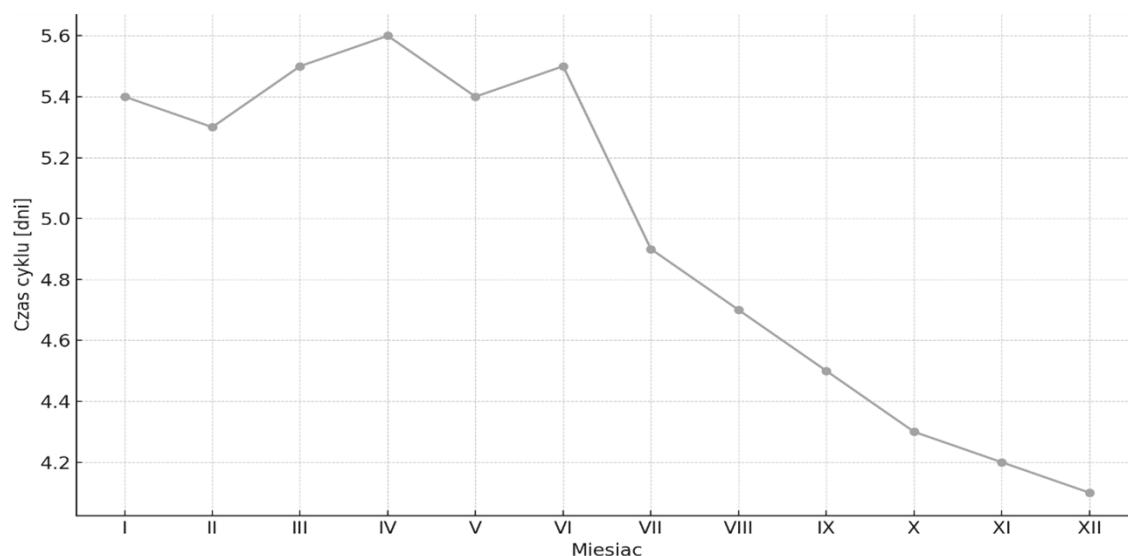
Klient jest zaangażowany w cykl PDCA: proces składania skarg zgodny z normą ISO 10002 systematycznie organizuje odbiór, klasyfikację, reakcję i ocenę skuteczności podjętych działań; wskaźniki dotyczące czasu, przejrzystości i bezstronności wzmacniają zaufanie i informują o bieżących ulepszeniach (Wierzowiecka, 2017).

Trwałość efektów zależy od dostosowanych ram organizacyjnych. Strukturyzacja działań wokół strumieni wartości, ustanowienie wyraźnych uprawnień decyzyjnych oraz wdrożenie krótkich ścieżek eskalacji ułatwiają przepływ informacji, zapewniając jednocześnie niezbędny nadzór. Kluczowe kwestie przy projektowaniu struktur obejmują podział pracy, hierarchię, centralizację lub decentralizację oraz wybór relacji - liniowych, funkcjonalnych, technicznych czy informacyjnych (Gembala, 2013).

Integracja BPM, PDCA, Lean i ISO jest skutecznie przedstawiona na poniższych wykresach. Na rysunku 1 widoczne jest skrócenie średniego czasu cyklu po wprowadzeniu limitów WIP i standaryzacji pracy. Stabilizacja obserwowana od 7.

miesiąca oznacza zmniejszenie zmienności i lepszą kontrolę nad wąskimi gardłami w kontroli jakości

Rysunek 1. Średni czas cyklu etykietowania – 12 miesięcy (przed/po)



Źródło: opracowanie własne na podstawie ISO 9001 i cykl PDCA (Kafel, Sikora, 2015)

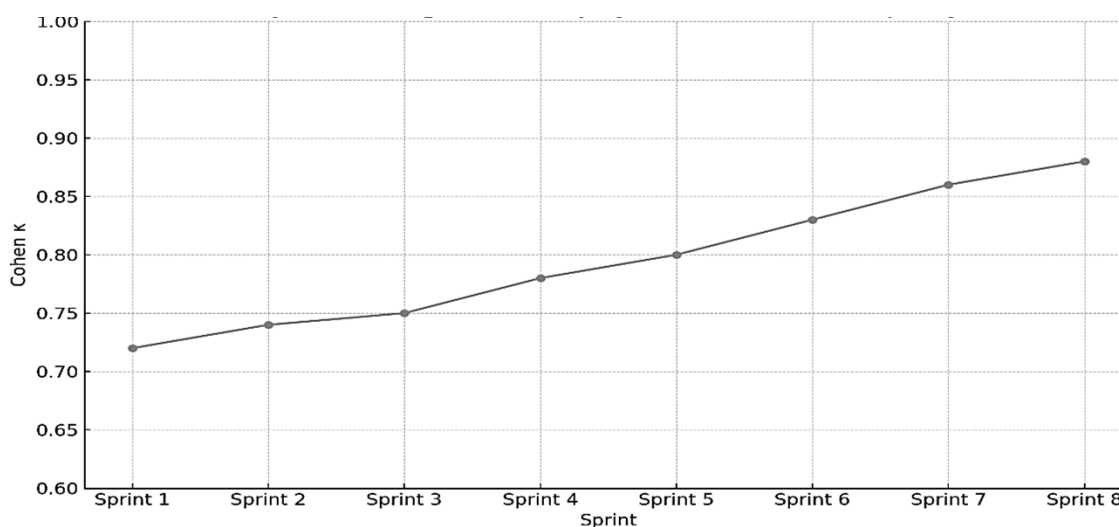
Rysunek 1 przedstawia dwunastomiesięczny przebieg średniego czasu cyklu etykietowania. W pierwszej połowie roku proces utrzymuje się na relatywnie stałym, lecz dość wysokim poziomie około 5,3–5,6 dnia. Od miesiąca siódmego widać wyraźną zmianę dynamiki: wartości zaczynają systematycznie maleć z 4,9 do 4,1 dnia na koniec okresu, co wskazuje na trwały efekt usprawnienia przepływu.

Różnica między okresem przed i po wdrożeniu rozwiązań jest wyraźna. Średnia dla miesięcy I–VI wynosi około 5,45 dnia, natomiast dla miesięcy VII–XII spada do około 4,45 dnia. Oznacza to skrócenie cyklu o pełny dzień, czyli redukcję rzędu 18,3% względem poziomu wyjściowego. Co ważne, w końcówce okresu - w miesiącach X–XII - przebieg stabilizuje się wokół 4,20 dnia przy bardzo małej zmienności (odchylenie standardowe rzędu 0,08 dnia), co sugeruje, że proces wszedł w fazę ustalonej pracy.

Interpretacyjnie taki kształt krzywej jest charakterystyczny dla skutecznego wdrożenia praktyk przepływu, takich jak limity WIP, standaryzacja pracy czy lepsze bramki jakości. Najpierw pojawia się skokowa poprawa wynikająca z usunięcia wąskich gardeł, a następnie efekt uczenia się organizacji, który „dociska” wynik w kolejnych miesiącach. Warto zauważyć, że pozorna większa rozpiętość wartości w okresie VII–IX wynika głównie z samego trendu spadkowego, a nie z rozchwiania procesu; dopiero płaski odcinek X–XII odsłania faktyczną, niską zmienność.

Zarządco wykres sugeruje, że przyjęte interwencje przyniosły nie tylko efekt szybkości, lecz także poprawę przewidywalności. Aby utrzymać ten stan i ewentualnie domknąć lukę do poziomu 4,0 dnia, warto kontynuować przeglądy PDCA, trzymać dyscyplinę limitów WIP i standardów pracy oraz monitorować dwa elementy wrażliwe na odbicie: obciążenie kontroli jakości i jakość wejść (kompletność oraz parametry obrazów). Takie podejście minimalizuje ryzyko powrotu do dłuższych cykli i pozwala przenieść ciężar działań z gaszenia odchyłeń na precyzyjne sterowanie stabilnym procesem.

Rysunek 2. Zgodność międzyadnotatorska (κ) - sprinty



Źródło: opracowanie własne na podstawie (Matych, Ligarski, 2015)

Rysunek 2 pokazuje stopniowy wzrost zgodności międzyadnotatorskiej (κ) w ośmiu kolejnych sprintach: od wartości 0,72 na początku do 0,88 w sprincie ósmym. Trajektoria ma dwie fazy: najpierw łagodną poprawę między sprintami 1–3 (0,72 \rightarrow 0,75), a następnie wyraźniejsze przyspieszenie od sprintu 4, gdzie krzywa „odrywa się” od poziomu wyjściowego i w sprintach 5–8 wchodzi w obszar wysokiej spójności (0,80 \rightarrow 0,88). Sumarycznie oznacza to bezwzględny przyrost κ o 0,16, czyli około 22% względem wartości bazowej, co w praktyce przekłada się na rzadsze rozbieżności interpretacyjne i mniejszą liczbę sporów wymagających rozstrzygnięcia (Matych, Ligarski, 2015).

Warto porównać średnie wartości dla początku i końca serii. Średnia z pierwszych trzech sprintów wynosi około 0,737, podczas gdy średnia z ostatnich trzech sprintów wzrasta do około 0,857. Takie przesunięcie wskazuje nie tylko na efekt pojedynczej interwencji, ale na utrwalanie się wspólnego modelu mentalnego zespołu i coraz lepsze

„docięcie” instrukcji adnotacyjnych do trudnych przypadków. Zaobserwowany skok w okolicach sprintów 4–5 jest spójny z wdrożeniami podwójnego znakowania i cyklicznych kalibracji, które zwykle działają jak katalizator: szybciej ujawniają reguły sporne i dostarczają materiału do precyzyjnych uzupełnień wytycznych.

Z perspektywy operacyjnej rosnąca κ ma podwójny efekt: ogranicza przepływ do adjudykacji (mniej przypadków spornych trafia „na górę”) i obniża rework, co skraca czas zamykania partii i stabilizuje przepływ w kontroli jakości. Charakterystyczne spłaszczenie krzywej między sprintami 7 i 8 (0,86 → 0,88) sugeruje zbliżanie się do górnej granicy możliwości przy danym profilu danych i poziomie szczegółowości wytycznych; dalsze przyrosty będą prawdopodobnie wymagały bardziej selektywnych interwencji, ukierunkowanych na konkretne klasy i typowe pomyłki, a nie wyłącznie na ogólną kalibrację (Matych, Ligarski, 2015).

Zarządco wykres wzmacnia tezę o skuteczności krótkich pętli uczenia się: iteracyjne przeglądy próbek, rejestr i analiza rozbieżności między adnotatorami, doprecyzowanie definicji klas oraz biblioteka przykładów brzegowych tworzą mechanizm, który najpierw podnosi κ , a potem pomaga ją utrzymać. Utrwalenie efektu wymaga dyscypliny w wersjonowaniu wytycznych i ciągłości kalibracji, ale także kontroli ryzyka „przeuczenia na instrukcję” - dlatego wzrost zgodności warto stale zestawiać z miarami jakości zewnętrznej (np. wpływem na skuteczność modelu) i monitorować, czy poprawa κ nie odbywa się kosztem nadmiernego uszczegóławiania zasad (Gruszka, 2018).

Podsumowując, hybryda BPM × PDCA/TQM × Lean × ISO × elementy agile zapewnia spójny łąd: architekturę procesów, rytm jakości, narzędzia przepływu, bezstronność i zarządzanie ryzykiem oraz iteracyjność zmian. Harmonizacja - zamiast wybiórczego wdrażania - przekłada się na przewidywalność, skalowalność i wiarygodność wyników, czyli kluczowe atrybuty konkurencyjności usług etykietowania. (Jasińska, 2017; Winiowski, 2016; Gruszka, 2018).

1.3. Problemy i wyzwania z zarządzaniem laboratorium etykietującym dane wizyjne

Główne wyzwania zarządcze w laboratorium etykietującym wynikają z charakteru danych wizyjnych: są heterogeniczne, wielomodalne i silnie zależne od kontekstu, co zwiększa niepewność interpretacyjną i podatność na rozbieżności między



adnotatorami. Bez systemowych mechanizmów integracji informacji, weryfikacji wiarygodności i redukcji niejednoznaczności łatwo dochodzi do utrwalania błędów zamiast ich wygaszania (Brodny, 2022).

Na etapie wejściowym procesu, zmienność warunków akwizycji obrazu – takich jak oświetlenie, optyka i czujniki - wraz z niekompletnymi metadanymi może stanowić poważne wyzwanie. Brak standaryzowanych parametrów obrazowania i nieregularne przetwarzanie wstępne zwiększają koszty interpretacji i prawdopodobieństwo konieczności przeróbek, co z kolei obniża jakość i czas cyklu. Wyniki te są zgodne z zaleceniami ustalonymi w praktyce inżynierskiej dla systemów wizyjnych wykorzystywanych w zastosowaniach przemysłowych (pełny przegląd rozwiązań można znaleźć w literaturze branżowej).

Kolejny szereg wyzwań dotyczy technologii i integracji. Laboratoria często korzystają z różnorodnych narzędzi i usług o różnym stopniu zaawansowania, którym towarzyszą zmienne wersje i interfejsy. W przypadku braku spójnej architektury integracyjnej, testów regresyjnych i dobrze zdefiniowanego mapowania danych pojawiają się problemy takie jak przestoje, utrata śladu audytu oraz komplikacje w wypełnianiu zobowiązań umownych i regulacyjnych. Problemy te są uznawane w literaturze poświęconej integracji systemów zarządzania za częste skutki rozdrobnienia odpowiedzialności i duplikacji procesów (Rajkiewicz i Mikulski, 2016; Kafel i Sikora, 2015).

Wymiar zgodności, prywatności i bezpieczeństwa rośnie wraz z upowszechnieniem treści zawierających dane wrażliwe. Materiał wizualny może ujawniać tożsamość lub okoliczności, których przetwarzanie wymaga szczególnych podstaw prawnych i środków minimalizacji. Włączenie obsługi skarg i reklamacji do pętli PDCA (np. przez procedury zbieżne z ISO 10002) skraca czas identyfikacji słabości procesu i dostarcza danych o jakości zewnętrznej, a w obszarach narażonych na manipulacje (deepfake) konieczne stają się dodatkowe punkty kontrolne i narzędzia detekcyjne (Chodyka, Ciekanski, 2024).

Wyzwania w strukturach społeczno-organizacyjnych zależą przede wszystkim od czynników związanych z kompetencjami i obciążeniem pracą. Dysproporcje jakości pogłębia monotonia zadań, przeciążenie poznawcze i wysoka rotacja personelu. W związku z tym zarówno literatura, jak i praktyka akredytowanych laboratoriów podkreślają konieczność jasno zdefiniowanego podziału ról, spójnych ewaluacji



merytorycznych i regularnych kalibracji zespołów – opartych na ustalonych wskaźnikach i ramach odpowiedzialności (Mytych, Ligarski, 2016).

Popyt na usługi etykietowania charakteryzuje się znaczną zmiennością, co przekłada się na presję zarówno na koszty, jak i wydajność. W przypadku braku ograniczeń związanych z produkcją w toku, precyzyjnego prognozowania i elastycznego zarządzania zasobami, rośnie prawdopodobieństwo nieprzestrzegania umów o poziomie usług (SLA), a także koszty związane z nadgodzinami lub nieuregulowanym outsourcingiem. Jednocześnie rośnie popularność wdrażania półautomatycznego wsparcia opartego na sztucznej inteligencji, takiego jak sugestie dotyczące etykiet. Chociaż zwiększa to produktywność, jednocześnie wprowadza nowe wymagania dotyczące zarządzania danymi, nadzoru i kompetencji cyfrowych (Doligalski, Kaszyński, 2025).

Szczególą kategorię stanowią wyzwania „jakości definicji”. Instrukcje etykietowania muszą równoważyć potrzeby modelu z ograniczeniami poznawczymi człowieka; brak przykładów brzegowych i jasnych reguł rozstrzygania sporów skutkuje lokalnymi interpretacjami i spadkiem spójności. Dobrą praktyką jest repozytorium wzorców, wersjonowanie księgi etykiet oraz obowiązkowe kalibracje - rozwiązania opisywane w literaturze zarządzania jakością jako filary stabilności kryteriów oceny (Kafel, Sikora, 2015).

Dynamicznie rośnie też znaczenie danych syntetycznych i treści manipulowanych. Bez walidacji źródeł i testów detekcyjnych ryzykujemy „zanieczyszczenie” zbiorów uczących i utratę wiarygodności wyników. Analizy z obszaru bezpieczeństwa państwa zwracają uwagę na potrzebę systemowych mechanizmów rozpoznawania manipulacji audiowizualnej, co można transferować do praktyk laboratoriów etykietujących (Chodyka, Ciekanski, 2024).

Wyzwanie związane z integracją różnych systemów zarządzania, takich jak zarządzanie jakością, bezpieczeństwem informacji i zarządzanie laboratorium, wciąż istnieje. Problem ten często rozwiązuje się poprzez tworzenie wspólnych map procesów, macierzy odpowiedzialności i jednolitych harmonogramów aktualizacji. Badania i literatura sugerują, że udane wdrożenia priorytetowo traktują ujednoczenie metryk i przepływów pracy, eliminując jednocześnie zbędną dokumentację, skracając w ten sposób czas zmian i minimalizując koszty utrzymania systemu (Rajkiewicz, Mikulski, 2016).



Podsumowując, efektywne zarządzanie laboratorium etykietującym wymaga zarówno elastyczności, jak i dyscypliny. Kluczowe jest, aby krótkie iteracje i kalibracje były zintegrowane z rygorystycznymi protokołami zgodności, wersjonowaniem wytycznych oraz regularnymi ocenami zarządzania. Ta połączona metodologia z jednej strony minimalizuje ryzyko nieporządku, a z drugiej zapobiega paraliżowi formalnemu, co bezpośrednio wpływa na niezawodność, terminowość i koszty związane z produkcją etykiet spełniających standardy jakości wymagane w projektach ML (Mytych, Ligarski, 2016; Kafel, Sikora, 2015; Rajkiewicz, Mikulski, 2016).

1.4. Analiza rynku i konkurencyjności laboratorium

Rynek etykietowania danych dynamicznie rośnie dzięki wdrażaniu sztucznej inteligencji (AI). Jednak przewaga konkurencyjna zależy obecnie nie tylko od kosztu jednostkowego, ale także od takich czynników, jak weryfikowalna jakość, skrócony czas realizacji, przestrzeganie norm regulacyjnych i bezpieczeństwo danych. W literaturze zwraca się uwagę, że tempo i skala wdrożeń AI bywają przeszacowywane przez szacunki ankietowe i przekaz marketingowy, dlatego decyzje zarządcze warto opierać na mierzalnych efektach i ryzykach, nie na hype-cyklach (Doligalski, Kaszyński, 2025).

Tabela 2. Mapa konkurencji – segmenty rynku i przykładowe przewagi

Segment rynku	Główne przewagi/dyferencjatory	Przykładowi gracze (charakterystyka)
Usługi (BPO/managed teams)	Skala zasobów, SLA i TAT, ścieżki QA, bezpieczeństwo i zgodność	Dostawcy globalni i regionalni; kontrakty SLA, elastyczność zasobów
Platformy komercyjne (SaaS/on-prem)	Auto-labeling, kontrola jakości w przepływie, integracje MLOps	Dostawcy narzędzi CV/NLP; API, plug-iny, analityka QA
Open-source (self-hosted)	Brak vendor lock-in, zgodność RODO/UE, on-prem	CVAT, Label Studio (wdrożenia własne / przez integratorów)
Integratorzy / software house (PL/UE)	Dostosowania, on-prem, łączenie z pipeline'ami klienta	Firmy lokalne łączące adnotację z integracją i utrzymaniem

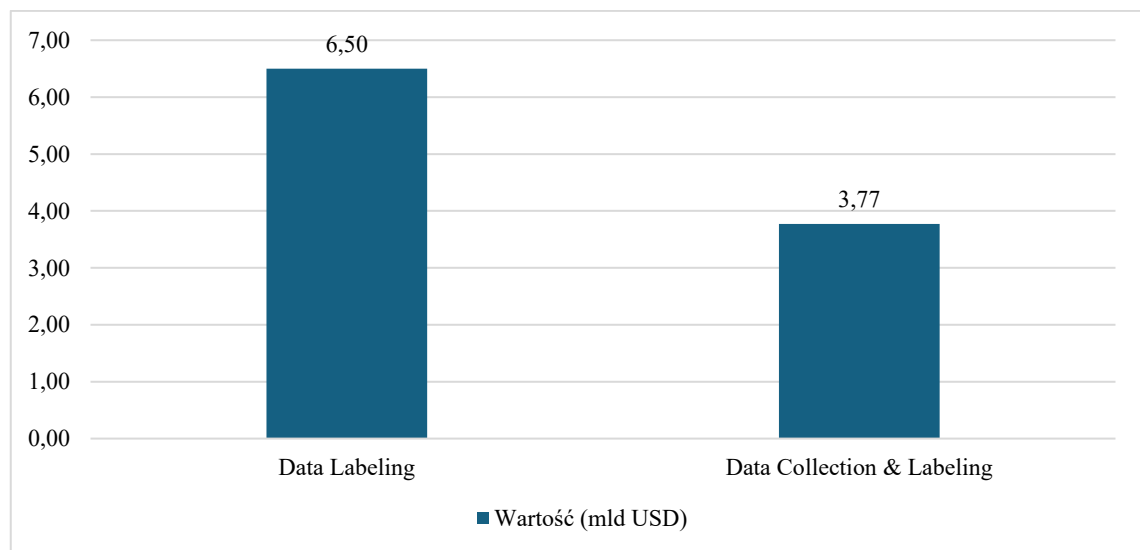
Źródło: opracowanie własne na podstawie przeglądu narzędzi SuperAnnotate (2025) oraz oficjalnych stron i dokumentacji wybranych platform (Labelbox, V7, Encord, AWS SageMaker Ground Truth, CVAT, Label Studio) i usługodawców (TELUS International AI Data Solutions).



Obecnie laboratoria usługowe konkurują ze sobą pod względem czasu potrzebnego na uzyskanie jakości (TAT, wraz z określonym poziomem zgodności κ i koniecznością poprawek) oraz zdolności do działania w ramach ram zgodności i bezpieczeństwa ustalonych przez klientów. Tę przewagę wzmacniają platformy i systemy OSS poprzez integrację funkcji półautomatycznych i MLOps. W związku z tym kluczowe jest połączenie kompetencji operacyjnych (SLA, QA) z możliwościami integracji po stronie laboratoryjnej.

W ujęciu regulacyjnym rynek Unii Europejskiej kładzie coraz większy nacisk na transparentność źródeł danych, dokumentowanie procesów oraz zapewnienie jakości treści wizualnych. Ponadto materiały audiowizualne wymagają zabezpieczeń przed manipulacją i dostępności narzędzi do wykrywania deepfake'ów, co skutkuje specyficznymi wymaganiami dla dostawców danych i etykiet (Chodyka, Ciekanski, 2024).

Wykres 1. Wielkość rynku (mld USD) wg wybranych raportów



Źródło: opracowanie własne na podstawie Grand View Research (2024), Data Collection, Labeling Market; Mordor Intelligence (2025), Data Labeling Market

Różnice wynikają z odmiennych parametrów definicyjnych; dla kierownictwa najważniejsze nie są liczby bezwzględne, lecz raczej ich implikacje: ciągły wzrost popytu, rosnące znaczenie audytowalności i zgodności oraz konieczność dostosowania się do ram MLOps.

W Polsce wysoka dojrzałość operacyjna sektora BSS/SSC/BPO, w połączeniu z możliwością realizacji projektów zgodnie z przepisami UE (takimi jak RODO i wymagania on-premises), buduje przewagę konkurencyjną. Tę przewagę dodatkowo wzmacnia stosunkowo niski poziom wdrożenia sztucznej inteligencji (AI) wśród krajowych przedsiębiorstw, co prowadzi do przesunięcia portfeli w kierunku eksportu usług. Dla laboratoriów taka sytuacja sprzyja naturalnej specjalizacji w obszarach takich jak zgodność i bezpieczeństwo danych, a także integracji narzędzi i wiedzy specjalistycznej.

Tabela 3. Rynek polski – czynniki podaży i popytu

Obszar	Wnioski dla laboratorium	Implikacje zarządcze
Podaż (BSS/SSC/BPO)	Skala i praktyki operacyjne dostępne na rynku pracy	Łatwiej budować SLA i TAT, szybsza skalowalność zasobów
Popyt krajowy	Adopcja AI < średnia UE (eksport usług)	Pozycjonowanie na rynki regulowane UE, języki/obszary
Zgodność/bezpieczeństwo	RODO, on-prem/sovereign EU, certyfikacje	„Compliance-first” jako element oferty i bariery wejścia
Ekosystem narzędziowy	SaaS, OSS + integratorzy	Unikaj lock-in; buduj zdolność integracyjną i QA-analitykę

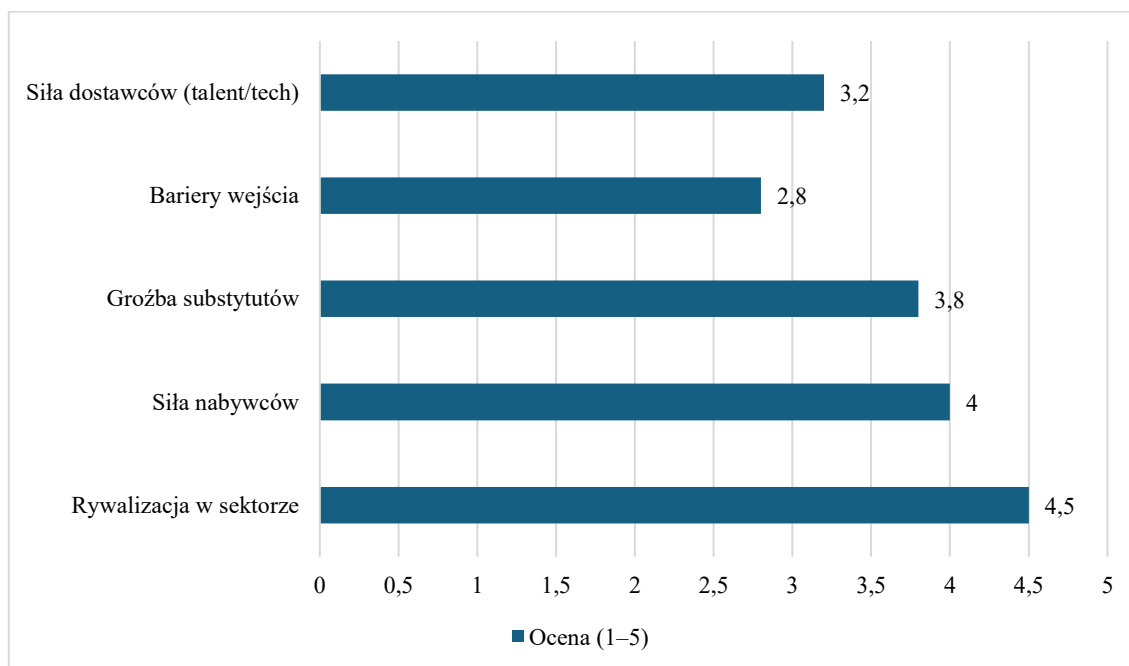
Zródło: opracowanie własne

Polska oferuje korzyści zarówno regulacyjne, jak i wykonawcze. Jednak ostatecznie sukces odniosą laboratoria, które z powodzeniem połączą kwestie jakości i zgodności z przepisami z wiedzą specjalistyczną w zakresie integracji i specjalizacją w określonych dziedzinach, takich jak technologia medyczna czy motoryzacja/pojazdy zaawansowane.

Poziom konkurencji w tym sektorze jest znaczny, charakteryzującym się dobrze poinformowanymi i wymagającymi nabywcami, podczas gdy bariery wejścia na rynek narzędzi są umiarkowane. W związku z tym, podejścia strategiczne muszą kłaść nacisk na różnicowanie poprzez jakość, zgodność z przepisami i specjalizację, a nie wyłącznie na cenę. To uzasadnienie oceny pozycjonowania konkurencyjnego znajduje potwierdzenie w modelu pięciu sił Portera, który jest omawiany w sposób syntetyczny w podręcznikach zarządzania (Porter, 2008).



Wykres 2. Pięć sił Portera – natężenie



Źródło: opracowanie własne na podstawie (Porter, 2008)

Obecność silnej konkurencji i znaczący wpływ nabywców wymuszają położenie nacisku na jakość i możliwość audytu; jednocześnie rozwój alternatywnych rozwiązań, takich jak automatyczne etykietowanie i dane syntetyczne, powoduje, że ludzie stają się teraz „ekspertami w temacie”, co należy brać pod uwagę przy opracowywaniu procesów i metryk.

Z punktu widzenia zarządzania i zgodności, wnioski z akredytowanych laboratoriów są istotne: spójna jakość opiera się na jasno określonym podziale obowiązków, regularnych ocenach, aktualizacjach wytycznych i ustalonych wskaźnikach kontroli (Mytych, Ligarski, 2016). Aby zapobiec nakładaniu się dokumentów i niejasności w zakresie odpowiedzialności, do integracji wymagań z różnych norm i systemów jakości należy podejść w sposób architektoniczny (Rajkiewicz, Mikulski, 2016).

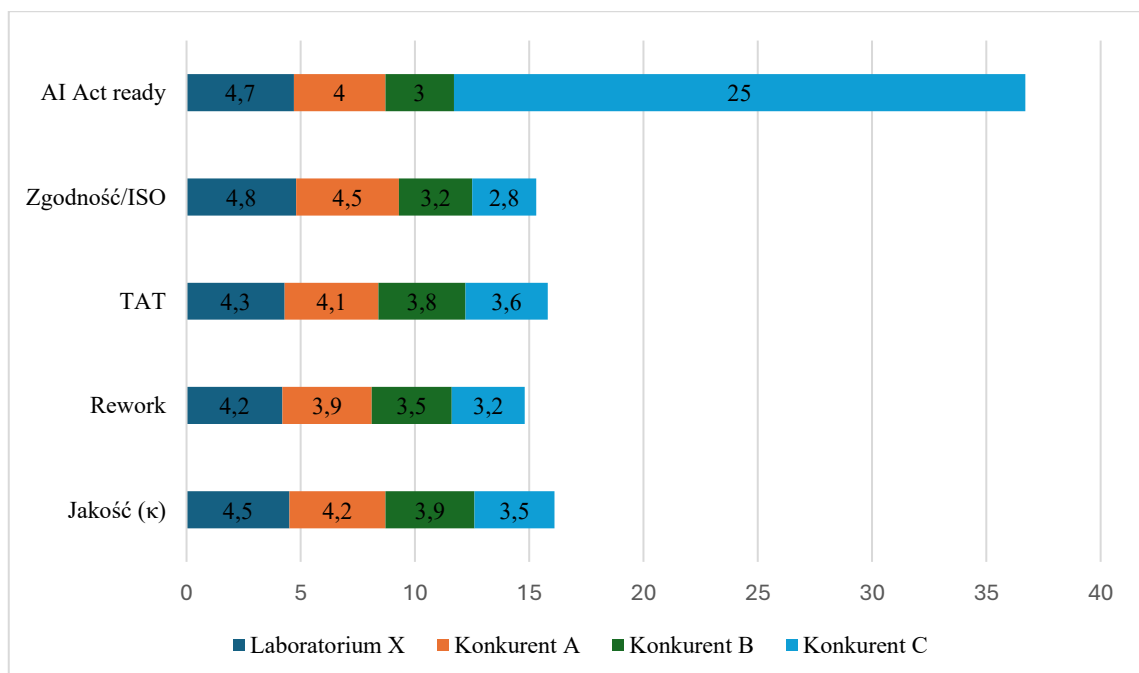
Tabela 4. Benchmark konkurencyjny

Kryterium	Laboratorium x	Konkurent A (usługi)	Konkurent B (SaaS)
Zgodność międzyadnotatorska κ (średnia)	0,86	0,83	0,80
Rework po I linii QA [%]	6,5	8,0	10,0
TAT – czas realizacji partii (dni)	4,0	4,5	5,0
Certyfikacje (ISO 9001/27001)	Tak	Tak	Nie
Gotowość do wymogów UE (ślad danych, Art. 10)	Wysoka	Średnia	Niska
Tryb dostawy (on-prem/sovereign EU)	Tak	Częściowo	Nie
Specjalizacja domenowa	Medtech; Automotive	Retail; General CV	General CV

Źródło: opracowanie własne na podstawie (Mytych, Ligarski, 2016)

Tabela 4 ilustruje, gdzie przewagę budują jakość i zgodność, a gdzie należy skracać TAT lub doskonalić kontrolę jakości (por. praktyki doskonalenia i nadzoru w laboratoriach akredytowanych).

Wykres 3. Benchmark punktowy (0–5) – atrybuty konkurencyjności



Źródło: opracowanie własne na podstawie (Rajkiewicz, Mikulski, 2016)



Wykres stanowi cenne narzędzie w rozmowach handlowych, podkreślając zarówno mocne strony – takie jak zgodność z normami ISO i możliwości lokalne - jak i identyfikując luki, które należy wyeliminować, takie jak konieczność dalszego skrócenia czasu realizacji zamówień bez uszczerbku dla jakości. System punktacji został opracowany zgodnie z zaleceniami, aby wyróżnić obszary o najwyższych barierach wejścia, w tym kompetencje, zgodność z przepisami i wiedzę specjalistyczną, zamiast koncentrować się na aspektach, które można łatwo powielić pod względem cen. Aby utrzymać przewagę konkurencyjną, niezbędne jest utrzymanie dyscypliny procesowej i integracja systemów jakości (Rajkiewicz, Mikulski, 2016).

Podsumowując rosnące zapotrzebowanie na wysokiej jakości etykiety, wraz z wymogami przejrzystości określonymi przez UE, podkreślają konieczność prowadzenia ścieżki audytu i mierzalnych wskaźników jakości; kluczowe jest odróżnienie decyzji inwestycyjnych od wpływu szumu informacyjnego (Doligalski, Kaszyński, 2025). W Polsce przewaga konkurencyjna powinna koncentrować się na podejściu „compliance-first” i płynnej integracji z procesami obsługi klienta, ponieważ eksport usług rekompensuje niższą adopcję sztucznej inteligencji (AI) na rynku krajowym. Najskuteczniejszym sposobem budowania pozycji konkurencyjnej jest: zapewnienie jakości (κ /rework), zarządzanie czasem realizacji (TAT) w ramach Statystycznej Kontroli Procesów (SPC), przestrzeganie standardów zgodności (ISO 9001/27001 + wymogi UE) oraz specjalizacja w określonych dziedzinach; ponadto integracja systemów i architektura procesów odgrywają kluczową rolę w minimalizacji kosztów koordynacji i ograniczaniu ryzyka wdrożeniowego (Mytych, Ligarski, 2016; Rajkiewicz, Mikulski, 2016).

Rozdział 2. Istota i przebieg procesu etykietowania danych wizyjnych

Etykietowanie danych wizyjnych (ang. *labeling, annotation*) rozumiane jest jako systematyczne przypisywanie obrazom informacji semantycznych klas, atrybutów i położeń (ramki, maski, punkty kluczowe) tworzących „prawdę referencyjną” dla uczenia nadzorowanego; w praktyce przemysłowej stanowi to warstwę opisową, na której opiera się dalsza automatyczna analiza i kontrola jakości wizyjnej (Połomski, 2016; B&R, iAutomatyka, et.al). Historycznie ewolucja tej warstwy następowała równolegle z dojrzewaniem systemów wizyjnych: od prostych czujników i kamer inteligentnych po rozwiązania wbudowane i aplikacje czasu rzeczywistego w robotyce, pojazdach autonomicznych i monitoringu (Gorgoń, et.al, 2013; Mielczarek, Muszyński, 2022). W strumieniu ML/AI etap etykietowania łączy akwizycję z trenowaniem i walidacją modeli, a jakość adnotacji współdecyduje o skuteczności algorytmów (dokładności, czułości) oraz o koszcie cyklu wytwórczego - co widać zwłaszcza w zastosowaniach, gdzie wizja zasila sterowanie i decyzje w czasie rzeczywistym (Mielczarek, Muszyński, 2022; Gorgoń, et.al, 2013).

Główne czynniki rozwoju obejmują postęp techniczny (oświetlenie, optyka, przetwarzanie w GPU/FPGA), profesjonalizację łańcucha dostaw danych i rosnący popyt na niezawodną analizę wizyjną w produkcji, bezpieczeństwie i automotive (B&R, iAutomatyka; Gorgoń, et.al, 2013). Jednocześnie ekspansja rozwiązań wizyjnych podnosi wagę ochrony prywatności i danych osobowych (PII/RODO), co wymusza łączenie środków organizacyjnych (polityki, retencja, oznaczanie stref) z technicznymi (anonimizacja odwracalna/nieodwracalna, kontrola dostępu, architektury bezpiecznego przepływu danych) - kompromis dokumentowany w literaturze zarówno na poziomie regulacyjnym, jak i algorytmicznym (Cichowski, Czyżewski, 2013).

Bilans wdrożeń wskazuje na przewagi: poprawę jakości i spójności, skrócenie czasu reakcji, ograniczenie kosztów błędów i reklamacji oraz wzrost przewidywalności procesu (Połomski, 2016; B&R, iAutomatyka), przy jednoczesnych barierach w postaci nakładów technologicznych i organizacyjnych, złożoności integracji oraz kosztów danych i zgodności - m.in. zwiększonej objętości strumieni/metadanych w architekturach prywatności (Cichowski, Czyżewski, 2013).

2.1. Założenia i specyfika procesu oznaczania danych

Etykietowanie danych wizyjnych to kontrolowane przypisywanie obrazom znaczeń od etykiety dla całego ujęcia, przez ramki i punkty kluczowe, po maski pikselowe w celu zbudowania wiarygodnej „prawdy referencyjnej” dla uczenia nadzorowanego oraz oceny modeli (B&R, iAutomatyka; Gorgoń et al., 2013). Jakość tej warstwy opisowej zależy równocześnie od dojrzałości procesu (jasne role, właściciele, KPI, przeglądy) i od warunków akwizycji (optyka, oświetlenie, parametry sensora), które determinują rozróżnialność cech i powtarzalność pracy anotatorów (Jasińska, 2017; Rajkiewicz, Mikulski, 2016; B&R, iAutomatyka). W podejściu procesowym zaleca się domkniętą pętlę PDCA i myślenie oparte na danych, co łączy praktykę etykietowania z wymaganiami ISO 9001 (Kafel, Sikora, 2015; Rajkiewicz, Mikulski, 2016).

W praktyce rozróżnia się cztery podstawowe zadania adnotacyjne. Klasyfikacja przypisuje etykietę całemu obrazowi i jest najtańsza, lecz najmniej informatywna przestrzennie (B&R, iAutomatyka).

Detekcja obiektów lokalizuje instancje klas za pomocą ramek (bbox), a jej jakość ocenia się m.in. przez IoU, krzywe precision–recall i AP/mAP (Osuchowski; Gorgoń et al., 2013). Segmentacja - semantyczna i instancyjna - opisuje granice na poziomie pikseli (maski), zapewniając najwyższą precyzję kosztem nakładu (B&R, iAutomatyka; Połomski, 2016). Keypoints kodują geometrię/pozę, szczególnie przydatną w robotyce i wizyjnym sprzężeniu zwrotnym (Mielczarek, Muszyński; Rodziewicz-Bielewicz, 2025).

Dobór formatów danych powinien być zgodny z pipeline’em ML/AI: dla detekcji dominują struktury JSON/XML z koordynatami ramek i klasą; dla segmentacji - maski binarne lub indeksowe; dla keypoints - listy punktów/szkieletów. Standaryzacja metadanych (słowniki kategorii, wersjonowanie wytycznych, kontrola zmian) obniża ryzyko błędów integracyjnych i stabilizuje walidację modeli (Gorgoń et al.; B&R, iAutomatyka). W praktyce rośnie udział pre-labelingu i aktywnego uczenia, które redukują koszt pozyskania etykiet przy zachowaniu jakości predykcji (Gorgoń et al., 2013).

Wytyczne anotacyjne muszą rozstrzygać przypadki brzegowe (occlusions, przezroczystości, odbicia, obiekty zagregowane), definiować zasady rysowania (jak blisko prowadzić maskę, jak liczyć instancje) i określać kryteria akceptacji. Spójność interpretacyjna jest wzmacniana przez hierarchię kontroli jakości: podwójne oznaczanie

klas ryzyka, przeglądy peer-review, adjudykację sporów i repozytorium wzorców; metryki jakości etykiet powinny korespondować z metrykami modeli (IoU, AP/mAP, F1) (Jasińska, 2017; Gorgoń et al., 2013).

Ergonomia i koszt procesu zależą także od projektu akwizycji: jednorodne oświetlenie, odpowiedni obiektowy i stabilizacja ujęcia wzmacniają cechy rozróżnialne przez anotatora, skracają czas rysowania i ograniczają rework (B&R, iAutomatyka; PAR - przegląd rozwiązań inspekcyjnych). W zastosowaniach zgodnych z GMP (np. farmacja) formalne kryteria kontroli nadruków i obecności etykiet (progi dla OCR/kodów) pokazują, jak łączyć wymagania procesu z wymaganiami metryk modeli (Połomski, 2016).

W obszarach zawierających dane osobowe (PII) konieczne są minimalizacja zakresu, anonimizacja (także odwracalna), kontrola dostępu i ślad audytowy przepływu danych; kompromis między użytecznością a prywatnością osiąga się, łącząc architekturę przepływu metadanych i algorytmy detekcji/zanonimizowania obszarów wrażliwych (Cichowski, Czyżewski, 2013).

Z perspektywy zarządzania procesowego dojrzałe etykietowanie wymaga jasnych ról i uprawnień (właściciele procesu, QA), PDCA, myślenia opartego na ryzyku i benchmarkingu wyników, co jest spójne z ewolucją od klasycznego TQM do współczesnego BPM (trzecia fala) - integrującego IT, automatyzację i ciągle ulepszanie (Jasińska, 2017; Rajkiewicz, Mikulski, 2016; Kafel, Sikora, 2023).

Tabela 5. Rodzaje zadań etykietowania, typowe formaty i metryki jakości

Zadanie	Jednostka adnotacji	Typowe formaty wyjściowe	Przykładowe metryki jakości	Zastosowania / uwagi	Względny koszt
Klasyfikacja	Etykieta dla całego obrazu	CSV/JSON (etykieta na obraz)	Precision, Recall, F1	Prescreening, kontrola zgodności partii	Niski
Detekcja obiektów	Ramki (x, y, w, h) + klasa	JSON/XML (bbox + klasa)	IoU, AP/mAP, Precision–Recall	Lokalizacja/liczenie, obecność/pozycja	Średni
Segmentacja (semant. /inst.)	Maska pikselowa / kontur	Maski binarne/indeksowe, COCO-like	mIoU, AP (segm.), Dice	Dokładne granice, defekty powierzchni	Wysoki



Keypoints	Zestawy punktów / szkielet	JSON (punkty/połączenia)	PCK, OKS, AP (keypoints)	Geometria/poza, montaż, robotyka	Średni-wysoki
-----------	----------------------------	--------------------------	--------------------------	----------------------------------	---------------

Źródło: opracowanie własne na podstawie przeglądów i instrukcji systemów wizyjnych (Gorgoń et al., 2013; B&R, iAutomatyka; Mielczarek, Muszyński), zastosowań QA w etykietowaniu (Połomski, 2016) oraz literatury o metrykach detekcji/segmentacji (Osuchowski, 2021).

Tabela 5 przedstawia, że im bogatsza jednostka adnotacji (od klasy → ramka → maska/punkty), tym większa informacyjność i potencjalnie lepsze metryki walidacyjne modeli; równocześnie rośnie koszt pozyskania etykiet, dlatego poziom szczegółowości należy dobierać do wymogów biznesowych i metryk akceptacyjnych, zamykając pętlę PDCA na danych i procesie (Jasińska, 2017; Rajkiewicz, Mikulski, 2016).

Podsumowując specyfika etykietowania danych wizyjnych sprowadza się do świadomego doboru poziomu szczegółowości adnotacji (klasyfikacja → detekcja → segmentacja → keypoints), który równoważy koszt pozyskania etykiet z wartością informacyjną dla modelu i metrykami walidacyjnymi (IoU, precision-recall, AP/mAP); jakość tej warstwy jest tak dobra, jak dobre są wytyczne anotacyjne i warunki akwizycji (optyka, oświetlenie, stabilność ujęcia), dlatego projekt stanowiska wizyjnego i standaryzacja formatów (JSON/XML dla bbox, maski indeksowe dla segmentacji, listy punktów dla keypoints) powinny być zaprojektowane „pod” docelowy pipeline ML/AI (Gorgoń et al., 2013; Osuchowski, 2021). Równocześnie, w zastosowaniach regulowanych - jak kontrola etykiet i nadruków - twarde kryteria akceptacji (np. progi czytelności dla OCR/kodów) muszą być zestrojone z metrykami modeli i procedurami QA, co obniża rework i całkowity koszt jakości (Połomski, 2016).

Tam, gdzie w materiałach mogą występować dane osobowe, nieodzowne są minimalizacja (krypto)anonimizacja i kontrola dostępu, wsparte architekturami bezpiecznego przepływu danych - kompromis między użytecznością a prywatnością potwierdza literatura przedmiotu (Cichowski, Czyżewski, 2013). Krótko: dojrzałe etykietowanie to połączenie inżynierii obrazu, jasnych wytycznych i modelocentrycznego pomiaru jakości, zamknięte w pętli PDCA i zgodne z wymaganiami zgodności - tylko wtedy dane uczące mają przewidywalną wartość dla pipeline’u ML/AI (Gorgoń et al.; Połomski, 2016).

2.2. Etykietowanie w ML/AI

Etykietowanie łączy inżynierię obrazu z etapem treningu i walidacji, tworząc „prawdę referencyjną” dla uczenia nadzorowanego i zamykając pętlę doskonalenia



w pipeline’ie ML/AI (planowanie danych → akwizycja → labeling → trening/walidacja → wdrożenie → monitoring/feedback). W sensie zarządczym to tutaj najłatwiej powiązać metryki jakości danych (zgodność międzyadnotatorska, rework) z metrykami modeli (IoU, AP/mAP) oraz wdrożyć mechanizmy zgodności i etyki przed propagacją błędów na kolejne etapy (Jasińska, 2017; Cichowski, Czyżewski, 2013).

2.2.1. Sztuczna inteligencja i uczenie maszynowe

W typowym strumieniu Computer Vision etykietowanie następuje po zdefiniowaniu celu i zaprojektowaniu toru akwizycji (kamera-optyka-oświetlenie), a przed treningiem oraz oceną modeli. Dane z etykietami dzieli się na zbiory treningowe/walidacyjne/testowe; po wdrożeniu obserwacje z eksploatacji (błędy, „długi ogon” klas, dryf danych) wracają do backlogu danych i labelingu w pętli PDCA lub w schemacie aktywnego uczenia (Jasińska, 2017; Osuchowski, 2021).

W warstwie metod ML etap etykietowania jest nośnikiem „nadzoru” - to tutaj do danych wstrzykuje się kontekst normatywny (pożądane rozróżnienia i wartości), co odróżnia uczenie nadzorowane od beznadzorowego i ogranicza utrwalanie niepożądanych wzorców z otwartych zbiorów (Horbowska, 2024).

Parametry akwizycji i pre-processingu (normalizacja, stabilizacja, segmentacja wstępna) warunkują rozróżnialność cech i pracochłonność adnotacji - poprawne przygotowanie toru wizyjnego upraszcza pracę anotatorów i stabilizuje trening (B&R, iAutomatyka; Podpora, 2011).

W sektorach przemysłowych pipeline spleciony jest z wymaganiami czasu rzeczywistego i ograniczeniami sprzętowymi (GPU/FPGA), co przenosi ciężar na jakość danych i etykiet już na wejściu (Gorgoń et al., 2013).

Gdy materiał zawiera PII, architektury bezpiecznego przepływu i anonimizacji należy zaprojektować przed etapem adnotacji, by zrównoważyć użyteczność zbioru i prywatność (Cichowski, Czyżewski, 2013).

Tabela 6. Miejsce etapu etykietowania w pipeline’ie ML/AI - etapy, artefakty, decyzje i KPI

Etap pipeline’u	Główne artefakty	Kluczowe decyzje	Przykładowe KPI
Definicja problemu	Cele biznesowe, kryteria sukcesu	Jakie klasy/pojęcia mają znaczenie decyzyjne?	Użyteczność metryk, SLA
Akwizycja i pre-processing	Kamera/optyka/światło, normalizacja	Parametry toru wizyjnego i przygotowania danych	Stabilność akwizycji,



			odsetek odrzuceń
Etykietowanie (labeling)	Guidelines, taksonomia, etykiety	Poziom szczegółowości (klasa/bbox/maska/kp), standard rysowania	κ (zgodność), rework, lead time/etykieta
Podział danych	Train/val/test	Proporcje, stratyfikacja, długi ogon	Równowaga klas, „leakage check”
Trening i walidacja	Model, hiperparametry	Dobór architektury/augmentacji	IoU, AP/mAP, F1, stabilność
Wdrożenie i monitoring	Model prod, logi, alerty	Progi decyzji, drift detection	TTA, odsetek błędów, drift danych/etykiety

Źródło: opracowanie własne na podstawie (Jasińska, 2017; Gorgoń et al., 2013; Osuchowski, 2021).

Etykietowanie jest „węzłem” pipeline’u - łączy definicje z akwizycją i warunkuje sensowność metryk modeli; PDCA domyka pętlę dzięki monitoringowi i powrotowi obserwacji do backlogu danych (Jasińska, 2017; Gorgoń et al. 2013; Osuchowski, 2021).

2.2.2. Rola etykietowania danych

Jakość i spójność etykiet wyznaczają górny pułap jakości modeli: systematyczne błędy (np. „ciasne/luźne” ramki, niespójne definicje klas) obniżają IoU, spłaszczają krzywe precision-recall i redukują AP/mAP; z kolei iteracyjne dosztacanie danych (aktywny dobór trudnych przykładów) i kalibracje anotatorów podnoszą metryki bez lawinowego wzrostu wolumenu (Osuchowski, 2021)

Rola labelingu jest także etyczno-prawna: na tym etapie łatwiej zidentyfikować i zneutralizować elementy wrażliwe (PII), zastosować anonimizację i kontrolę dostępu oraz zapewnić zgodność z wytycznymi dla systemów wysokiego ryzyka (Cichowski, Czyżewski, 2013; Horbowska, 2024).

Tabela 7. Błędy etykiet → wpływ na metryki modeli → diagnostyka i działania korygujące

Typ błędu etykiety	Wpływ na metryki modelu	Symptomy diagnostyczne	Działania korygujące
Ramki „za ciasne/za luźne”	Spadek IoU; AP/mAP↓	Zawyżone FP/FN na granicach	Rekalibracja zasad rysowania, przykłady wzorcowe, QA 2-linii
Niespójna definicja klas	Precision/Recall↓, mylenie klas	Wysoki rework, niska κ	Doprecyzowanie taksonomii, przykłady edge-case, szkolenia



Brak trudnych etykiet	AP na długim ogonie↓	Błędy w rzadkich klasach	Active learning/hard-example mining, oversampling
Błędy masek (dziury, „schody”)	mIoU↓, artefakty na krawędziach	Niska zgodność konturu vs GT	Reguły wygładzania, maski z piórem, adjudkacja
Dryf wytycznych	Niestabilne PR-curves	„Skoki” jakości po sprintach	Wersjonowanie wytycznych, checklisty, retraining

Źródło: opracowanie własne na podstawie (Osuchowski, 2021; Jasińska, 2017; Gorgoń et al., 2013).

Największe straty w AP/mAP wynikają z błędów systematycznych i dryfu wytycznych; ich wykrywanie opiera się na łączeniu metryk modelu z metrykami procesu (κ , rework) i audytem partii danych (Osuchowski, 2021; Jasińska, 2017).

Tabela 8. Strategie pozyskiwania etykiet a koszt i jakość

Strategia	Idea	Efekt na koszt/tempo	Efekt na jakość/metryki	Kiedy stosować
Pre-labeling	Wstępne podpowiedzi modelu	Czas/etykietę↓, throughput↑	Jakość = funkcja QA i korekt	Gdy mamy model bazowy i jasne wytyczne
Active learning	Dobór „najbardziej informacyjnych” próbek	Liczba etykiet↓ (celowana)	AP/mAP↑ na trudnych przypadkach	Długi ogon klas, ograniczony budżet
Double-label + adjudkacja	Podwójne oznaczanie i rozstrzygnięcie sporów	Koszt↑	κ ↑, mniej błędów systematycznych	Klasy/ryzyka krytyczne
Kalibracja okresowa	Sesje z przykładami edge-case	Krótkotrwałe koszt↑	Stabilność metryk↑, rework↓	Po zmianie danych/narzędzi/rotacji

Źródło: opracowanie własne na podstawie (Osuchowski, 2021; Jasińska, 2017; Gorgoń et al., 2013)

Pre-labeling i active learning ograniczają liczbę etykiet bez utraty jakości, o ile towarzyszy im kontrola QA; double-label opłaca się dla klas wysokiego ryzyka, bo podnosi κ i stabilizuje AP (Gorgoń et al., 2013; Osuchowski, 2021).



Tabela 9. Prywatność i zgodność w pipeline’ie CV – kontrolki na etapie etykietowania

Ryzyko/obszar	Kontrolka na etapie labelingu	Przykłady praktyk	Uwagi zgodności
PII w obrazie	Minimalizacja/anonimizacja (odwr. /nieodwr.)	Zaciemnianie twarzy/tablic, pseudonimizacja	Privacy by design, rejestr czynności
Dostęp do materiału	Least-privilege, VPN, konta korporacyjne	Role-based access, logowanie zdarzeń	Dowody audytowe, retencja
Przepływ danych	Praca w środowisku klienta	Brak kopiowania poza VPN	Umowy powierzenia, DPIA
Etyka danych	Zasady wykluczania treści szkodliwych	Repo przypadków zabronionych	Procedury eskalacji

Źródło: opracowanie własne na podstawie (Cichowski, Czyżewski, 2013)

Najtaniej i najbezpieczniej neutralizować ryzyka prywatności przed adnotacją: minimalizacja, anonimizacja i ścisłe role dostępu zmniejszają koszty późniejszych korekt i ryzyka prawne (Cichowski, Czyżewski, 2013).

Podsumowując etykietowanie to węzeł strategiczny pipeline’u: decyduje o tym, jakie sygnały trafią do modelu i jak będą one mierzone; jego jakość (definicje, wytyczne, QA) wprost wyznacza pułap metryk walidacyjnych, a jego organizacja (PDCA/BPM, kontrolki prywatności) minimalizuje ryzyka techniczne i regulacyjne (Jasińska, 2017; Osuchowski, 2021; Cichowski, Czyżewski, 2013; Horbowska, 2024).

2.3. Wyzwania i oczekiwania wobec nowoczesnych modeli

Nowoczesne modele CV (w tym „foundation models”) są coraz bardziej dane-i etykieto-chłonne, ale jednocześnie rośnie oczekiwanie efektywności danych (więcej jakości z mniejszej liczby etykiet), odporności (na szum etykiet i przesunięcia domeny), ogólności (zero-/few-shot), śledzalności (datasheets/model cards) i zgodności regulacyjnej (EU AI Act - art. 10). Poniżej syntetyzuję kluczowe wyzwania i praktyki, które powinny kształtować organizację labelingu.

Długi ogon klas i nierównowaga. W danych z rzeczywistej produkcji dominują klasy „główne”, podczas gdy klasy rzadkie są dramatycznie niedoreprezentowane, co



pogarsza uogólnianie modeli. Aktualne przeglądy wskazują, że skuteczne podejścia łączą rekonstrukcję reprezentacji, re-ważenie strat i projektowanie zadań/augmentacji pod ogon klas (oraz świadome próbkowanie do adnotacji). To problem powszechny i słabo „domknięty” nawet w dużych benchmarkach. (Yang, Jiang, Song, Guo, 2022; Fu et al., 2022).

Efektywność pozyskiwania etykiet (AL, pre-labeling). W praktyce rośnie rola active learning i pre-labelingu: selekcja „najbardziej informacyjnych” próbek oraz podpowiedzi modeli skracają czas etykietowania bez utraty jakości - o ile towarzyszy im rygor QA i kalibracje anotatorów. Stan badań obejmuje zarówno ujęcia ogólne, jak i wyspecjalizowane (np. obrazowanie medyczne). (Li et al., 2024; Wang et al., 2024; Wan et al., 2023).

Szum i błędy etykiet. Niespójności definicji klas, ramki „za ciasne/za luźne” czy artefakty masek obniżają IoU i AP/mAP; stąd oczekiwanie na robust training (np. techniki filtracji/ważenia przykładów, korekta etykiet, loss-correction) oraz audyty partii danych przed treningiem. Aktualne przeglądy opisują spektrum metod radzenia sobie z szumem w CV i obrazowaniu medycznym. (Song et al., 2020; Shi et al., 2024; Wei et al., 2024).

Modele „promptowalne” i automatyzacja adnotacji. Foundation models, np. Segment Anything (SAM), obiecują znaczne obniżenie kosztu przygotowania masek, ale ich działanie bywa dalekie od „doskonałości” - specyfika domen (tekstury, okluzje, 3D) ujawnia ograniczenia i konieczność człowieka-w-pętli. Ujęcia „SAM-w-domenie” (np. MedSAM) poprawiają transfer, ale nie eliminują wymogu QA. (Kirillov et al., 2023; Ma et al., 2023; Ji et al., 2023; przegląd SAM 2023).

Odporność na przesunięcia domeny i drifty. Wdrożenia oczekują stabilności przy zmianie warunków (oświetlenie, optyka, tło, sezonowość), co wymaga ciągłego uczenia (active/continual learning), walidacji pod-dystrybucyjnej oraz polityk aktualizacji zbiorów referencyjnych i wytycznych adnotacyjnych - z jasnym śladem audytowym (datasheets/model cards). (Geburu et al. 2013, 2018; Mitchell et al., 2019).

Śledzalność i dokumentacja danych/modeli. Oczekiwany dziś standard to Datasheets for Datasets i Model Cards - dokumenty, które wiążą parametry zbioru (pochodzenie, ograniczenia, dozwolone użycia) oraz modelu (metryki w różnych warunkach) z procesem decyzyjnym i ryzykiem. W praktyce ułatwiają zgodność, re-trening i transfer wiedzy w organizacji. (Geburu et al. 2013, 2018/2021; Mitchell et al., 2019).



Zgodność regulacyjna (EU AI Act). W systemach wysokiego ryzyka art. 10 nakazuje stosowne praktyki data governance dla zbiorów treningowych/walidacyjnych/testowych (pochodzenie, celowość, czyszczenie, adekwatność). Horyzont stosowania przepisów jest określony (wejście w życie 1.08.2024; kolejne obowiązki do 2026/2027), więc organizacje oczekują, by pipeline danych i etykietowania był „compliance-ready” (ślad pochodzenia, retencja, kontrola dostępu). (EUR-Lex 2024/1689; AI Act - art. 10).

Tabela 10. Wyzwania nowoczesnych modeli CV a oczekiwane praktyki danych/etykietowania

Obszar wyzwania	Objaw w metrykach	Implikacje dla labelingu	Oczekiwane praktyki
Długi ogon klas	Spadek AP na klasach rzadkich	Potrzebne ukierunkowane etykiety dla ogona	AL/oversampling, re-ważenie, projekt klas
Szum etykiet	IoU↓, AP/mAP↓, κ↓	QA i filtry etykiet przed treningiem	Robust training, loss-correction, audyt partii
Koszt etykiet	Budżet/lead time↑	Automatyzacja „z człowiekiem-w-pętli”	Pre-labeling, AL, kalibracje
Foundation models	Zmienna jakość zero-shot	Weryfikowalność i korekty masek	QA 2-stopnia, domenowe dostrajanie
Drift/domena	Niestabilność PR-curves	Aktualizacja zbiorów i wytycznych	Continual learning, datasheets/model cards
Zgodność (AI Act)	Ryzyko prawne	Udokumentowane pochodzenie danych	Data governance, kontrola dostępu, DPIA

Źródło: opracowanie własne na podstawie Yang, Jiang, Song Guo (2022); Li, Li, Ren (2024); Song, Kim, Park, Shin, Lee (2020); Wan, Zhang, Tan (2023); Kirillov et al. (2023); Ke et al. (2023); Gebru et al., 2013; Mitchell et al. (2019); Regulation (EU) 2024/1689 (AI Act), art. 10.

Tabela 10 zestawia najczęstsze punkty krytyczne współczesnych modeli CV z praktykami, jakie powinny ukierunkowywać organizację etykietowania. W pierwszym wierszu „długi ogon klas” podkreśla, że większość danych reprezentuje klasy „główne”, a rzadkie wystąpienia budują lwią część błędów stąd oczekiwanie, by etykietowanie było selektywne, a nie równomierne: aktywne uczenie i przemyślana taksonomia kierują budżet w „ogon”, gdzie każdy nowy przykład daje największy przyrost AP. Przeglądy długiego ogona pokazują, że właśnie dobór próbek i re-ważenie są bardziej skuteczne niż samo zwiększanie wolumenu danych bez strategii.



W drugim wierszu „szum etykiet” akcentuje się, że jakość ground truth wyznacza górny pułap metryk modelu. Literatura wskazuje dwutorowe podejście:

- kontrola jakości przed treningiem (audyt partii danych, podwójne oznaczanie klas ryzyka, adjudykacja),
- robust training po stronie modeli (ważenie przykładów, korekta strat).
Obie osie trzeba zaprojektować wspólnie, inaczej poprawki po stronie modelu jedynie maskują braki w danych.

Trzeci wiersz dotyczy kosztu etykiet. Oczekiwana dziś „efektywność danych” oznacza więcej jakości z mniejszej liczby adnotacji: pre-labeling skraca czas jednostkowy, a aktywne uczenie zmniejsza liczbę niezbędnych przykładów - pod warunkiem sprawnego QA i regularnych kalibracji anotatorów. Przeglądy pokazują, że dobrze zaprojektowane pętle AL osiągają wyższe zyski właśnie w obszarach „ogonu” i trudnych przypadkach.

Czwarty wiersz („foundation models/SAM”) wskazuje na realne, ale nieautomagiczne oszczędności. Modele promptowalne, jak Segment Anything, potrafią radykalnie przyspieszyć przygotowanie masek, lecz ich jakość jest zmienna domenowo; wymaga to człowieka-w-pętli i drugiego stopnia QA, a w zastosowaniach specjalistycznych - dostrajania do danej dziedziny.

Piąty wiersz („drift/domena”) przypomina, że warunki zbierania danych zmieniają się w czasie (sprzęt, oświetlenie, sezonowość), co rozstraja modele. Oczekiwana odpowiedź to ciągłe uczenie i udokumentowana ścieżka danych: datasheets i model cards porządkują pochodzenie, ograniczenia oraz metryki w różnych warunkach, co ułatwia świadome aktualizacje wytycznych adnotacyjnych.

Ostatni wiersz („zgodność (AI Act)”) łączy praktyki danych z wymaganiami prawnymi UE. Artykuł 10 rozporządzenia 2024/1689 wymaga zarządzania zbiorem (pochodzenie, adekwatność, czyszczenie, kontrola dostępu) dla treningu/walidacji/testów, przy harmonogramie wdrożenia rozłożonym do 2026/2027. Stąd oczekiwanie, by proces etykietowania był „compliance-ready”: mieć ślad pochodzenia i decyzji, zasady dostępu oraz retencję.

Podsumowując nowoczesne modele wizji komputerowej wymagają mądrego zarządzania etykietowaniem, a nie tylko większej liczby etykiet. Największe przyrosty jakości (AP/mAP, stabilność PR-curves) osiąga się przez ukierunkowane etykietowanie ogona klas, redukcję szumu etykiet (QA + robust training), automatyzację z człowiekiem-w-pętli (pre-labeling, SAM z QA 2-stopnia) oraz ciągłą dokumentację i monitorowanie



(datasheets/model cards), spójne z wymaganiami EU AI Act. Operacyjnie oznacza to „modelocentryczne” KPI dla etykiet (IoU/AP na próbkach kontrolnych, κ, rework), wersjonowane wytyczne adnotacji i architekturę zgodności (pochodzenie, dostęp, retencja). Tak zorganizowany proces etykietowania staje się dźwignią jakości i zgodności, a nie wąskim gardłem projektu ML/AI.

2.4. Ochrona danych i polityka prywatności (RODO/PII, anonimizacja, kontrola dostępu)

Etykietowanie danych wizyjnych nierzadko obejmuje informacje umożliwiające identyfikację osób (PII) — twarze, sylwetki, tablice rejestracyjne, znaki szczególne - dlatego proces musi być zaprojektowany zgodnie z zasadami RODO (legalność, ograniczenie celu, minimalizacja, prawidłowość, ograniczenie przechowywania, integralność i poufność oraz rozliczalność) oraz w duchu privacy by design/by default (art. 25) (European Union, 2016).

Wytyczne Europejskiej Rady Ochrony Danych dotyczące urządzeń wizyjnych precyzują test niezbędności, oznakowanie monitoringu, retencję i minimalizację (EDPB, 2020), a starsza, wciąż cytowana opinia Grupy Art. 29 pokazuje ograniczenia technik anonimizacji i ryzyko reidentyfikacji przy łączeniu zbiorów (WP29, 2014).

W obszarach wysokiego ryzyka dodatkowe wymagania wobec ładu danych (pochodzenie, adekwatność, czyszczenie, kontrola dostępu) ustanawia unijny AI Act (art. 10) (European Union, 2024).

Na poziomie systemowym rolę „szkieletu” dla kontrolek dostępu i dowodów rozliczalności pełnią normy ISO/IEC 27001 (ISMS) i ISO/IEC 27701 (PIMS) (ISO/IEC, 2022; ISO/IEC, 2019).

W kontekście wideo badania przeglądowe opisują również architektury i algorytmy ochrony prywatności (maskowanie selektywne, anonimizacja odwracalna/nieodwracalna, kontrola przepływu), które można bezpośrednio odwzorować na praktykę laboratorium etykietującego (Cichowski, Czyżewski, 2013).

Aby przełożyć wymagania prawne na codzienność laboratorium, warto „rozplaszczyc” ogólne zasady RODO do poziomu konkretnych ról, czynności i artefaktów audytowych. Wytyczne EDPB 3/2019 dotyczące wideo doprecyzowują niezbędność, oznakowanie i retencję, zaś WP29 (2014) przypomina, że sama anonimizacja nie zawsze gwarantuje nieidentyfikowalność. Uzupełnieniem są



systemowe standardy ISO/IEC 27001/27701, które zamieniają zasady w mierzalne kontrolki i dowody (EDPB, 2020; WP29, 2014; ISO/IEC, 2022; ISO/IEC, 2019).

Tabela 11. Zasady RODO → praktyki w laboratorium etykietującym → dowody zgodności

Zasada RODO	Co oznacza w labelingu CV	Praktyki operacyjne	Dowody/artefakty
Legalność i ograniczenie celu	Jasny, określony cel przetwarzania	Rejestr czynności, klauzule informacyjne/test proporcjonalności	Rejestr (art. 30), DPIA, noty celu
Minimalizacja danych	Tylko niezbędne dane i metryki	Kadrowanie, przycinanie klatek, wyłączanie audio; ograniczenie klas	Specyfikacja danych, backlog odrzuceń
Poufność i integralność	Ochrona materiału i etykiet	RBAC/ABAC, MFA, VPN, szyfrowanie w spoczynku i tranzycie	Matryca uprawnień, logi (SIEM)
Ograniczenie przechowywania	Retencja adekwatna do celu/ryzyka	Automatyczne kasowanie/archiwizacja po walidacji	Polityka retencji, raporty kasowania
Rozliczalność	Możliwość wykazania zgodności	Wersjonowanie wytycznych, ślad zmian etykiet, audyty partii	Datasheet zbioru, protokoły QA

Źródło: opracowanie własne na podstawie RODO (European Union, 2016), EDPB 3/2019 (EDPB, 2020), WP29 (2014) oraz ISO/IEC 27001/27701 (ISO/IEC, 2022; ISO/IEC, 2019).

Zestawienie pokazuje, jak każda zasada RODO materializuje się w praktyce labelingu: „legalność i cel” → rejestr czynności i DPIA; „minimalizacja” → cięcia kadrów i ograniczenie klas; „poufność” → RBAC/MFA/VPN i szyfrowanie; „retencja” → twarde reguły przechowywania; „rozliczalność” → wersjonowanie wytycznych i ślad zmian etykiet. Kluczowe jest projektowanie dowodów zgodności równolegle z procesem adnotacji (EDPB, 2020; WP29, 2014).

Dobór techniki ochrony PII to kompromis między siłą ochrony a użytecznością zbioru dla uczenia. EDPB i WP29 podkreślają, że rozmycie/pikselizacja przy niskich poziomach mogą być niewystarczające i wymagają oceny ryzyka reidentyfikacji; ENISA porządkuje praktyki pseudonimizacji i modele zagrożeń (EDPB, 2020; WP29, 2014; ENISA, 2019).



Tabela 12. Techniki ukrywania/ograniczania PII w obrazie - moc, ryzyko, zastosowania

Technika	Opis	Moc ochrony / ryzyko	Zastosowania typowe	Uwagi praktyczne
Maskowanie/wycięcie ROI	Zakrycie/usunięcie fragmentu	Wysoka ochrona, utrata kontekstu	Twarze, tablice, logotypy	Proste i skuteczne, ale może utrudniać kontekst
Rozmycie/pikselizacja	Filtr na obszarach PII	Średnia ochrona (ryzyko odzysku)	Monitoring, ulica	Wymaga kalibracji i testów skuteczności
Pseudonimizacja metadanych	Zamiana ID na losowe	Chroni tożsamości, nie obraz	Przepływy z ID zdarzeń	Należy chronić klucz mapowania
Anonimizacja odwracalna	„Zaszyte” maski do odsłonięcia	Wysoka kontrola, zależna od kluczy	Materiały dowodowe, audyt	Klucze poza zasięgiem adnotatorów
Anonimizacja nieodwracalna	Trwałe usunięcie cech	Najwyższa ochrona	Dane publiczne, repozytoria	Test reidentyfikacji + ocena użyteczności
Generowanie syntetyczne	Zamiana PII na dane syntetyczne	Zależna od jakości generatora	Rzadkie scenariusze	Ryzyko artefaktów i biasu syntetycznego

Źródło: opracowanie własne na podstawie EDPB 3/2019 (EDPB, 2020), WP29 (2014) oraz ENISA (2019).

Największą ochronę daje maskowanie/wycięcie (kosztem kontekstu), podczas gdy rozmycie/pikselizacja wymagają testów skuteczności. Pseudonimizacja zabezpiecza identyfikatory, ale nie obraz; gdy potrzebne jest warunkowe ujawnienie, sprawdza się anonimizacja odwracalna (z kontrolą kluczy). W praktyce laboratorium należy dobrać technikę do ryzyka i celu, a następnie sprawdzić wpływ na metryki modeli na próbach kontrolnych (EDPB, 2020; WP29, 2014; ENISA, 2019).

Ponieważ etykietowanie operuje na materiałach potencjalnie wrażliwych, kontrolki dostępu i dowody ich działania muszą być tak samo mierzalne jak metryki



jakości etykiet. ISO/IEC 27001/27701 rekomendują RBAC/ABAC, MFA, szyfrowanie i rejestry logów; AI Act art. 10 wzmacnia wymóg śladu pochodzenia danych dla zbiorów treningowych/walidacyjnych/testowych (ISO/IEC, 2022; ISO/IEC, 2019; European Union, 2024).

Tabela 13. Kontrola dostępu i ślad audytowy – matryca „zasób-zagrozenie-kontrola-dowód”

Zasób	Zagrozenie	Kontrola	Dowód/Ew. metryka
Materiał źródłowy (wideo/obrazy)	Nieuprawniony wgląd/eksfiltracja	RBAC/least privilege, VPN/VDI, szyfrowanie, zakaz lokalnych kopii	Logi dostępu, testy DLP, przeglądy uprawnień
Etykiety i metadane	Modyfikacja/wyciek	Repo z historią zmian, podpisy/hashe, backup	Audit trail, wyniki kontroli integralności
Środowisko pracy	Przejęcie konta	MFA, rotacja haseł, klucze sprzętowe	Raporty MFA, statystyki blokad
Transfer poza EOG	Nielegalny transfer	SCC/umowy powierzenia, szyfrowanie end-to-end	Rejestr transferów, DPIA, rejestr decyzji

Źródło: opracowanie własne na podstawie ISO/IEC 27001:2022, ISO/IEC 27701 oraz AI Act (Regulation (EU) 2024/1689, art. 10) (ISO/IEC, 2022; ISO/IEC, 2019; European Union, 2024).

Skuteczność ochrony nie wynika z samej polityki, lecz z logów, przeglądów uprawnień, testów DLP, podpisów integralności i dowodów retencji. Operacyjnie oznacza to pracę „w środowisku klienta” (VPN/VDI, zero-copy), zasady least privilege i podpisane łańcuchy zmian etykiet - spójnie z ISO/IEC i AI Act (ISO/IEC, 2022; ISO/IEC, 2019; European Union, 2024).

Skuteczna ochrona danych w procesie etykietowania to połączenie zasad RODO (celowość, minimalizacja, rozliczalność), technik ochrony obrazu (maskowanie, anonimizacja, pseudonimizacja) i twardych kontrolek dostępu (RBAC/MFA/VPN, logi, retencja), uzupełnione dokumentacją danych (datasheets, rejestry czynności) i DPIA. Wymogi te są spójne z oczekiwaniami AI Act wobec data governance i można je operacjonalizować przez ISO/IEC 27001/27701. W praktyce laboratorium najtaniej i najbezpieczniej neutralizować ryzyka przed adnotacją (privacy-by-design), a następnie utrzymywać ślad audytowy i mierniki zgodności na równi z metrykami jakości danych -



tak, by etykiety były nie tylko poprawne merytorycznie, ale także bezpieczne i zgodne (EDPB, 2020; WP29, 2014; European Union, 2016, 2024; ISO/IEC, 2019, 2022; ENISA, 2019; Cichowski, Czyżewski, 2013).

2.5. Definicja etykietowania/labelingu/annotation oraz etymologia

W literaturze przedmiotu trzy pojęcia - etykietowanie (pol. „etykieta” jako nalepka/oznaczenie), nadawanie etykiet/kategorii (ang. *labeling*) i dodawanie adnotacji/not (ang. *annotation*) - opisują nadawanie danym znaczenia poprzez przypisanie klas, atrybutów lub metadanych; semantycznie wywodzą się z odmiennych tradycji: „*label*” z dawnego francuskiego *label/labeau* („wąska tasiemka”) i franko-germańskich rdzeni oznaczających „płat/klapę” „*annotation*” z łac. *annotare* („zaznaczać, notować”) (Etymonline, 2025), zaś „etykieta” z fr. *étiquette* („bilet, karteczka”, potem „norma zachowania”). W informatyce i uczeniu maszynowym „*label*” oznacza zmienną docelową w uczeniu nadzorowanym, natomiast „*annotation*” zwykle obejmuje szersze metadane (np. ramki, maski, punkty kluczowe, relacje), które umożliwiają trening i walidację modeli (Goodfellow, Bengio, Courville, 2016).

Historycznie pojęcie *annotation* najwcześniej upowszechniło się w lingwistyce korpusowej (Penn Treebank: adnotacje POS i składni od lat 90.), stanowiąc wzorzec jakości i organizacji pracy adnotatorów (Marcus, Santorini, Marcinkiewicz, 1993; Taylor, 2003). W wizji komputerowej skokowe zwiększenie skali i precyzji adnotacji przyniosły PASCAL VOC (klasyfikacja/detekcja), ImageNet (masowe etykietowanie klas obiektów) oraz MS COCO (segmentacje instancji w kontekście), co umocniło dzisiejsze znaczenie *labelingu/annotation* jako „prawdy referencyjnej” w ML (Everingham et al., 2010; Deng et al., 2009; Lin et al., 2014). Jednocześnie w praktyce przemysłowej „etykietowanie” bywa rozumiane dosłownie - jako nanoszenie etykiet na wyroby i weryfikacja nadruków/kodów przez systemy wizyjne - co pokazuje różnorodność kontekstów i potrzeb standaryzacji terminów (Połomski, 2016; PAR, 2013).

Aby uniknąć niejednoznaczności między „etykieta” jako nalepką a „etykieta” jako kategorią klasyfikacji, poniższe zestawienie rozróżnia zakresy znaczeniowe i typowe artefakty w ML/AI, odwołując się do źródeł etymologicznych oraz klasycznych opracowań z CV/NLP (Goodfellow et al., 2016; Marcus et al., 1993; Everingham et al., 2010; Lin et al., 2014).



Tabela 14. Terminologia „etykietowanie / labeling / annotation” - znaczenie i użycie w ML/AI

Termin	Rdzeń i etymologia	Definicja robocza (IT/ML)	Typowe artefakty
Etykietowanie (PL)	fr. étiquette → pol. „etykieta” (nalepka/oznaczenie)	Nadawanie danym znaczenia poprzez przypisanie klas/stanów; w przemyśle także fizyczne naklejanie etykiet	Klasy, kody, nadruki; w CV: klasy bbox/maski
Labeling (EN)	OF label „tasiemka, znacznik” → „oznaczyć”	Przypisanie etykiety docelowej w uczeniu nadzorowanym	„y/target”, klasy, atrybuty
Annotation (EN)	łac. annotare „zaznaczać, notować”	Dodanie strukturalnych adnotacji/metadanych	Ramki, maski, keypoints, relacje, tagi

Źródło: Źródło: opracowanie własne na podstawie Goodfellow et al. (2016), Marcus et al. (1993), Everingham et al. (2010), Lin et al. (2014).

Interpretacja. W ML „labeling” najczęściej oznacza nadanie klasy/targetu, a „annotation” - szczegółowe odwzorowanie struktury obiektu (ramka, maska, punkty). W środowisku produkcyjnym warto jawnie wskazywać, czy mowa o „etykietowaniu danych” (ML) czy „etykietowaniu wyrobu” (jakościowo-logistycznym), aby uniknąć różnoznacności terminów (Połomski, 2016).

Upowszechnienie pojęć labeling/annotation w informatyce i AI przebiegało etapami: od lingwistyki korpusowej (standaryzowane adnotacje), przez benchmarki CV (VOC, ImageNet, COCO), po współczesne narzędzia półautomatyczne; kamienie milowe poniżej (Marcus et al., 1993; Everingham et al., 2010; Deng et al., 2009; Lin et al., 2014).

Tabela 15. Kamienie milowe upowszechnienia „annotation/labeling” w IT/AI

Rok	Dziedzina	Wydarzenie	Znaczenie dla praktyki labelingu
1993	NLP	Penn Treebank	Standaryzacja schematów adnotacji i organizacji pracy adnotatorów
2005–2010	CV	PASCAL VOC	Zestaw zadań (klasyfikacja/detekcja/segmentacja) i procedury ewaluacji



2009	CV	ImageNet	Skala etykietowania; crowdsourcing (AMT) i taksonomie klas
2014	CV	MS COCO	Segmentacje instancji „w kontekście” i wzorce adnotacji wysokiej precyzji

Źródło: opracowanie własne na podstawie Marcus et al. (1993); Deng et al. (2009); Lin et al. (2014).

Skuteczność ochrony nie wynika z samej polityki, lecz z logów, przeglądów uprawnień, testów DLP, podpisów integralności i dowodów retencji. Operacyjnie oznacza to pracę „w środowisku klienta” (VPN/VDI, zero-copy), zasady least privilege i podpisane łańcuchy zmian etykiet - spójnie z ISO/IEC i AI Act (ISO/IEC, 2022; ISO/IEC, 2019; European Union, 2024).

Skuteczna ochrona danych w procesie etykietowania to połączenie zasad RODO (celowość, minimalizacja, rozliczalność), technik ochrony obrazu (maskowanie, anonimizacja, pseudonimizacja) i twardych kontrolek dostępu (RBAC/MFA/VPN, logi, retencja), uzupełnione dokumentacją danych (datasheets, rejestry czynności) i DPIA. Wymogi te są spójne z oczekiwaniami AI Act wobec data governance i można je operacjonalizować przez ISO/IEC 27001/27701. W praktyce laboratorium najtaniej i najbezpieczniej neutralizować ryzyka przed adnotacją (privacy-by-design), a następnie utrzymywać ślad audytowy i mierniki zgodności na równi z metrykami jakości danych - tak, by etykiety były nie tylko poprawne merytorycznie, ale także bezpieczne i zgodne (EDPB, 2020; WP29, 2014; European Union, 2016, 2024; ISO/IEC, 2019, 2022; ENISA, 2019; Cichowski, Czyżewski, 2013)

2.6. Historia rozwoju etykietowania (kamienie milowe, narzędzia, standaryzacja QA)

Początki etykietowania danych wizyjnych wyrastały z tradycji przetwarzania obrazów i rozpoznawania wzorców: adnotacje wykonywano ad hoc na niewielkich zbiorach, aby móc oceniać algorytmy segmentacji, detekcji czy klasyfikacji. W polskim piśmiennictwie ugruntowanie tej bazy technicznej – kamery, odwzorowanie obrazu w pamięci, segmentacja progowa i metody jakości obrazu – dobrze dokumentują podręczniki dotyczące przetwarzania obrazów przemysłowych, które tworzyły kontekst dla późniejszej standaryzacji adnotacji i kontroli jakości danych (Rafajłowicz, Rafajłowicz, 2012).



Pierwszym przełomem w samym „labelingu” była popularyzacja internetowych narzędzi do wspólnej adnotacji. LabelMe zaproponowało już w połowie lat 2000. prosty interfejs rysowania wielokątów i udostępniania adnotacji, ustanawiając w praktyce konwencje formatów (XML/JSON), kategorii obiektów i bazodanowego przechowywania metadanych (Russell, Torralba, Murphy, Freeman, 2008).

Równolegle konkurs PASCAL VOC (2005–2012) przyniósł drugi trwały kamień milowy: spójny zbiór obrazów z jednolicie opracowanymi etykietami (głównie ramki ograniczające) oraz standardowe miary oceny (m.in. mAP), co pozwoliło porównywać metody na wspólnej skali i wymusiło bardziej rygorystyczne procedury adnotacyjne i weryfikacyjne (Everingham, Eslami, Van Gool, Williams, Winn, Zisserman, 2010).

Następna fala to skala: ImageNet, budowany przy użyciu platform crowdsourcingowych, pokazał, jak operacjonalizować masowe etykietowanie rozproszonych danych i jak walidować spójność kategorii w ontologii (WordNet). Ta zmiana skali wymusiła wielostopniowe przeglądy jakości i metody „golden setów” oraz walidacje krzyżowe, które trwale weszły do praktyki QA (Deng, Dong, Socher, Li, Li, Fei-Fei, 2009).

Kolejne rozszerzenie zakresu to MS COCO - przejście z samych ramek do segmentacji instancji i bogatszego opisu kontekstu sceny. Wymagało to nowych wytycznych adnotacyjnych (precyzja konturów, rozdzielanie nakładających się obiektów), a także wieloetapowych przeglądów jakościowych z miernikami zgodności między adnotatorami (Lin, Maire, Belongie, Hays, Perona, Ramanan, Dollár, Zitnick, 2014).

W tym czasie dojrzywał również krajowy kontekst zastosowań. Przemysłowe systemy wizyjne w Polsce zaczęły wykorzystywać etykietowanie nie tylko do badań, ale do inspekcji produkcji, OCR i weryfikacji nadruków, co pośrednio standaryzowało praktyki adnotacyjne (zbiory wzorców, słowniki klas defektów) i protokoły QA na styku produkcji i informatyki (Pomiary Automatyka Robotyka, 2013).

Równolegle opisywane wdrożenia w przemyśle farmaceutycznym – z naciskiem na GMP, zgodność zapisu i kontrolę poprawności etykiet – ilustrują, jak wymagania regulacyjne przyspieszyły formalizację procesów etykietowania i ich audytowalność (Połomski, 2016).

Od strony narzędziowej nastąpiła profesjonalizacja warsztatu adnotatora: lekkie, przeglądarkowe środowiska do adnotacji (VGG Image Annotator - VIA) obniżyły barierę wejścia i wprowadziły eksport do otwartych formatów, a otwartoźródłowe platformy



takie jak CVAT i Label Studio dołożyły skalowalność, współbieżną pracę zespołów, auto-annotację (np. wstępne maski/ramki) i API do integracji z pipeline'ami MLOps (Dutta, Zisserman, 2019).

Z dojrzewaniem narzędzi i skali rosły wymagania jakościowe. W praktykach ML upowszechniały się „datasheets for datasets” i „data statements” – metody dokumentowania pochodzenia danych, kryteriów włączenia/wyłączenia, ról adnotatorów, miar zgodności i ograniczeń użycia. Te ramy stały się de facto standardem meta-QA dla zbiorów danych, pomagając ograniczać błędy systematyczne i poprawiać replikowalność (Bender, Friedman, 2018).

Równolegle do praktyk branżowych rozwijały się formalne normy jakości danych, które - choć nie powstały z myślą o wizji komputerowej – dostarczyły wspólnego słownika cech jakości (np. kompletność, spójność, wiarygodność) i procedur oceny, co można adaptować do zbiorów adnotacji obrazowych (ISO/IEC 25012:2008; w danych przestrzennych - ISO 19157 i jego nowsze części). To zwróciło uwagę na mierzalność jakości etykiet i raportowanie wyników oceny w procesach odbioru danych (ISO/IEC 25012, 2008; ISO 19157, 2013/2023).

W Polsce badania i wdrożenia prac zespołów uczelni technicznej (m.in. przeglądy zastosowań systemów wizyjnych, od pojazdów autonomicznych po medycynę) dokumentują przejście od rozwiązań punktowych do złożonych łańcuchów przetwarzania, w których adnotacja jest powiązana z akceleracją obliczeń (GPU/FPGA), kontrolą strumieni wideo i reżimami oceny jakości, co przekłada się na rosnącą rolę standaryzacji danych uczących (Gorgoń i in., 2022).

Podsumowując, historia etykietowania danych wizyjnych to droga od ręcznie tworzonych, lokalnych notatek do złożonych, skalowalnych ekosystemów adnotacyjnych ze sformalizowanym QA: LabelMe wprowadziło wspólnotowe narzędzia online, VOC ustandaryzowało benchmarki i metryki, ImageNet zademonstrował skalę i walidację crowdsourcingu, COCO rozszerzyło zakres na segmentacje instancji, a współczesne platformy (VIA, CVAT, Label Studio) i ramy dokumentacyjne (datasheets/data statements) osadziły jakość danych w dojrzałych praktykach inżynierskich – także w Polsce, gdzie naciski regulacyjne i przemysłowe przyspieszyły adopcję metodyk QA w labeling-u produkcyjnym.

2.7. Czynniki wpływu i konieczność rozwoju

Rozwój procesu etykietowania danych wizyjnych przyspiesza wraz z postępem technologicznym i rosnącymi oczekiwaniami biznesowymi. Usprawnienia w przepustowości torów akwizycji oraz dostępności akceleracji obliczeń otwierają drogę do realizacji bardziej złożonych zadań widzenia maszynowego - od segmentacji instancji po analizę scen wieloobektowych. W tych warunkach kluczowe stają się standardy opisu i kontroli geometrii etykiet, które spajają inżynierię obrazu z etapem trenowania i walidacji modeli. W polskiej literaturze technicznej widać, jak ewolucja architektury systemów wizyjnych przekłada się na rosnące wymagania wobec jakości i spójności adnotacji: im większa rozdzielczość, dynamika sceny i różnorodność domen, tym większa rola standardów opisu i kontroli geometrii etykiet, które stają się w praktyce „łącznikiem” między inżynierią obrazu a skutecznym trenowaniem modeli (Gorgoń i in., 2022; Chełuska, 2018).

W tym otoczeniu przewagę uzyskują zespoły, które przekształciły etykietowanie z czynności pomocniczej w zorganizowany przepływ wartości danych. Współczesny „łańcuch dostaw danych” obejmuje pozyskanie materiału, definicję taksonomii, adnotację i weryfikację, wersjonowanie i retencję, a wreszcie publikację zestawów treningowych, walidacyjnych i testowych do środowisk MLOps. Polskie opracowania strategiczne i branżowe podkreślają, że inwestycje w jakość danych i procesy ich wytwarzania są warunkiem uprzemysłowienia AI - od deklarowanych celów (Polityki dla rozwoju SI w Polsce od 2020 r.) obejmujące kompetencje, infrastrukturę i otwarte zasoby danych), po raporty wdrożeniowe dla MŚP wskazujące konkretne bariery i wzorce wsparcia przy adopcji rozwiązań AI (KPRM, 2020; PARP, 2024; PARP, 2025).

Przesunięcie ciężaru na „dane jako produkt” sprzyja podejściu data-centric, w którym kolejne iteracje projektu koncentrują się na specyfikacjach etykiet, rewizjach wytycznych i audytach partii danych przed trenowaniem, a nie tylko na doborze architektury. Ten kierunek dobrze współbrzmi z polską tradycją „trzeciej fali” zarządzania procesami, kładącą nacisk na ciągłe usprawnianie, jawność zasad i mierzalność wyników - co w obszarze adnotacji przekłada się na formalne bramki jakości (np. progi IoU, zgodność międzyanotatorska) oraz wersjonowanie wytycznych i „złoty próbek” (Jasińska, 2017).

Równolegle rośnie presja regulacyjna i ładu danych. Sztandarowy akt prawny Unii Europejskiej - AI Act - ustanawia wymogi ładu danych dla systemów wysokiego ryzyka, wprost obejmujące pochodzenie, adekwatność, czyszczenie i nadzór nad



zbiorami treningowymi, walidacyjnymi i testowymi. Z perspektywy laboratoriów etykietujących oznacza to konieczność dokumentowania procesów adnotacyjnych, ról i uprawnień, kontroli dostępu oraz retencji, a także śladu decyzyjnego w doborze klas i metryk jakości, co scala dotychczas rozproszone praktyki w spójny system zarządzania danymi (European Union, 2024).

Na polskim rynku zapotrzebowanie na wiarygodne, dobrze udokumentowane zbiory rośnie wraz z upowszechnieniem rozwiązań AI. Badania NASK dotyczące wpływu gen-AI na rynek pracy pokazują szeroki zakres zawodów i zadań podatnych na automatyzację i wspomaganie, co pośrednio zwiększa popyt na zespoły i narzędzia wytwarzające dane uczące wysokiej jakości; jednocześnie raporty PARP wskazują, że w MŚP zasoby kompetencyjne i procesowe są wciąż nierównomierne, dlatego standaryzacja etykietowania i QA staje się elementem przewagi operacyjnej i redukcji ryzyka wdrożeń (NASK, 2025; PARP, 2025).

Wymienione wektory zmian znajdują odzwierciedlenie także w literaturze technicznej i praktyce produkcyjnej. Publikacje polskich ośrodków akademickich i branżowych prezentują dojrzewanie narzędzi oraz metod pracy zespołów adnotacyjnych - od definiowania klas i standardów rysowania konturów po organizację pracy wieloetapowej z niezależną weryfikacją - co ostatecznie zacieśnia sprzężenie zwrotne między kontrolą jakości danych a stabilnością metryk modeli w środowisku wdrożeniowym (Gorgoń, et al, 2022; Chełuska, 2018).

Konkludując, konieczność rozwoju procesu etykietowania wynika ze skali i złożoności współczesnych systemów wizyjnych, z rosnących wymagań „łańcucha dostaw danych” oraz z normatywnych oczekiwań wobec ładu danych w AI. Organizacje, które traktują etykietowanie jako pełnoprawny proces zarządczy - z właścicielami, miernikami, politykami jakości i zgodności - zyskują zdolność do szybszego reagowania na zmiany taksonomii, stabilniejszego dowożenia wartości i ograniczenia kosztów poprawek. W polskich realiach tę transformację wspierają dokumenty strategiczne państwa i programy wdrożeniowe, które podnoszą dostępność infrastruktury i kompetencji, a jednocześnie wzmacniają oczekiwania dotyczące transparentności i audytowalności danych uczących (KPRM, 2020; PARP, 2024; European Union, 2024).

2.8. Zalety

Profesjonalnie zorganizowane etykietowanie danych wizyjnych staje się źródłem przewagi w całym cyklu tworzenia rozwiązań ML/AI, bo łączy myślenie procesowe



(właściciele, KPI, PDCA) z metrykami modelocentrycznymi (IoU, AP/mAP), co przekłada się na powtarzalność wyników i skrócenie czasu iteracji decyzyjnych (Jasińska, 2017).

Z perspektywy konkurencyjności kluczowa jest jakość i szybkość „łańcucha dostaw danych”: stabilne wytyczne adnotacyjne, kontrola ryzyka etykiet i audytowalność zbiorów umożliwiają szybsze wprowadzanie modeli na produkcję oraz ich bezpieczne doskonalenie; polskie opracowania o systemach wizyjnych potwierdzają, że to dane i ich standardy, a nie wyłącznie architektury modeli, determinują realną wydajność wdrożeń w przemyśle (Gorgoń i in., 2022).

Korzyści operacyjne i kosztowe wynikają z redukcji reworku oraz lepszego dopasowania metryk QA do celów modelu: spójne ramki i maski, podwójne etykietowanie klas ryzyka oraz jasne progi akceptacyjne ograniczają błędy systematyczne, podnoszą IoU/AP i stabilizują krzywe precision-recall; doświadczenia z aplikacji wizyjnych w kontroli nadruków/etykiet dowodzą, że standaryzacja procedur etykietowania przekłada się na mniejszą liczbę poprawek i niższy całkowity koszt jakości (Osuchowski, 2021; Połomski, 2016).

Silne zaplecze etykietowania uruchamia też projekty rozwojowe: iteracyjne doprecyznianie taksonomii, aktywne uczenie i kalibracje anotatorów pozwalają efektywnie adresować „długi ogon” klas oraz wdrażać coraz bardziej złożone zadania CV (segmentacje instancji, keypoints) bez lawinowego wzrostu budżetu danych; w podejściu procesowym takie „uczenie się na danych” jest naturalnym elementem doskonalenia strumienia wartości (Jasińska, 2017; Gorgoń i in., 2013).

Wymiar współpracy sektorowej ujawnia się w ekosystemach uczelnia–przemysł: jasne standardy adnotacyjne, ślad audytowy i reguły jakości umożliwiają bezpieczny przepływ danych między partnerami, skracają onboarding zespołów i ułatwiają replikację wyników; zarządzanie jakością podkreśla, że standaryzacja i transparentność procesu są warunkiem trwałej kooperacji międzyorganizacyjnej oraz skalowania przedsięwzięć (Kafel, Sikora, 2015).

Na poziomie pozycji rynkowej przewagą stają się wiarygodność i zgodność: laboratorium, które potrafi wykazać pochodzenie zbiorów, kontrolę dostępu, retencję oraz skuteczne techniki minimalizacji/anonimizacji, buduje zaufanie klientów i skraca ścieżkę zgodności prawnej; polskie prace o ochronie prywatności w systemach wizyjnych oraz standardy dla CCTV wskazują praktyczne architektury i wymagania integralności zapisu,



które wzmacniają „compliance readiness” procesów etykietowania (Cichowski, Czyżewski, 2013).

Podsumowując, dobrze ułożone etykietowanie zwiększa konkurencyjność (szybsze i przewidywalne wdrożenia), poprawia efektywność (mniej poprawek, wyższe metryki modeli), otwiera projekty rozwojowe (trudniejsze zadania przy kontrolowanym koszcie danych), wzmacnia współpracę sektorową (wspólne standardy i dowody jakości) oraz pozycję rynkową (zgodność, bezpieczeństwo, zaufanie). W realiach polskich te korzyści są dodatkowo wspierane przez rosnące wykorzystanie systemów wizyjnych w przemyśle i jasno opisane ramy jakości/bezpieczeństwa danych.

2.9. Wady i ograniczenia / bariery

Etykietowanie danych wizyjnych, choć strategicznie niezbędne dla jakości modeli ML/AI, obarczone jest zestawem kosztownych i trudnych do wyeliminowania barier. Najbardziej oczywisty wymiar dotyczy nakładów technologicznych: nowoczesne stanowiska do pracy z obrazem i wideo wymagają wysokowydajnych stacji roboczych/GPU, macierzy pamięci oraz licencji na oprogramowanie adnotacyjne i systemy kontroli jakości. W literaturze technicznej dotyczącej systemów wizyjnych podkreśla się, że wraz ze wzrostem rozdzielczości i dynamiki sceny rośnie zapotrzebowanie na przetwarzanie i składowanie, co istotnie podnosi koszty infrastruktury dla całego łańcucha danych (Gorgoń i in., 2022).

Drugą grupę barier tworzą koszty zgodności i bezpieczeństwa, które narastają, gdy materiały zawierają dane osobowe (twarze, tablice rejestracyjne, kontekst miejsca). Opracowania o ochronie prywatności w systemach wizyjnych wskazują, że wdrożenie anonimizacji/pseudonimizacji generuje dodatkowe obciążenia obliczeniowe oraz zwiększa wolumen danych wyjściowych (np. skompresowane strumienie po przekształceniach), co bezpośrednio przekłada się na koszty infrastruktury i utrzymania (Cichowski, Czyżewski, 2013). Jednocześnie wybór algorytmów i architektur ochrony prywatności jest ograniczony przez dostępne zasoby obliczeniowe, co powoduje kompromisy między poziomem ochrony a wydajnością procesu (Cichowski, Czyżewski, 2013).

Wymogi normalizacyjne także nie są neutralne kosztowo. Prace przeglądowe dotyczące CCTV podkreślają, że projektowanie i eksploatacja systemów z zachowaniem norm PN-EN 62676 wiąże się z dodatkowymi wymaganiami dla jakości zapisu, integralności i archiwizacji, a zatem z inwestycjami w urządzenia, oprogramowanie

i procedury - co wprost odnosi się do laboratoriów pracujących na materiałach wizyjnych (Matuszek, 2020). Koszty te są trudne do „odchudzenia”, bo wynikają z konieczności zapewnienia powtarzalności i walidowalności danych oraz odporności systemu na incydenty (Matuszek, 2020).

Na poziomie zarządzania jakością i akredytacji dochodzą obciążenia dokumentacyjne i procesowe. Badania nad funkcjonowaniem akredytowanych laboratoriów zwracają uwagę, że utrzymywanie systemu (audyt wewnętrzny, nadzór nad dokumentacją, kwalifikacja personelu, raportowanie niezgodności) generuje stałe koszty operacyjne i wymaga dedykowanych kompetencji, co może ograniczać elastyczność i podnosić próg wejścia dla nowych jednostek (Mytych, Ligarski, 2016). W praktyce oznacza to, że nawet przy identycznym wolumenie danych koszt jednostkowy „etykiety wysokiej jakości” bywa istotnie większy w środowisku objętym rygorami akredytacyjnymi.

W literaturze zarządzania procesami podkreśla się dodatkowo ryzyko organizacyjne i transformacyjne: wdrożenia nowych metodyk pracy (np. przejście na iteracyjny, mierzalny workflow danych) potrafią się przeciągać, a zbyt radykalna zmiana bez przygotowania zespołu zwiększa prawdopodobieństwo porażki i kosztów ukrytych (Jasińska, 2017). Z perspektywy laboratorium etykietującego oznacza to konieczność łączenia wymogów jakości i zgodności z uważnym planowaniem zmian - w przeciwnym razie rosną koszty koordynacji, szkolenia i „odpadu” prac (Jasińska, 2017).

Kolejny wymiar to długi czas zwrotu z inwestycji. W obszarze systemów wizyjnych ROI zależy nie tylko od wdrożonego modelu, ale od całego łańcucha: pozyskanie/odtworzenie danych, ich etykietowanie, walidacja, integracja z produkcją i monitoring po wdrożeniu. Opracowania branżowe pokazują, że rynek reaguje na presję kosztową inwestycjami w „inteligentne” oprogramowanie, ale zysk pojawia się po osiągnięciu pewnej skali i stabilności procesu - co w praktyce może oznaczać wielomiesięczny horyzont zwrotu (Matuszek, 2020). W tym okresie organizacje finansują równocześnie dane, narzędzia, QA i eksperymenty modelowe.

Istotną barierą bywa brak dofinansowań dopasowanych do specyfiki prac nad danymi. Programy wsparcia częściej finansują „produkt końcowy” (model, wdrożenie) niż przygotowanie i utrzymanie danych uczących. W efekcie laboratoria ponoszą wysokie koszty bieżące (kompletowanie zbiorów, kontrola jakości, wersjonowanie, retencja), które trudno zrekompensować w ramach krótkich budżetów projektowych - zwłaszcza



gdy wymagane są stałe przeglądy i audyty zgodności (Mytych, Ligarski, 2016; Jasińska, 2017).

Na poziomie operacyjnym narzut kosztowy generują praktyki kontroli jakości: podwójne etykietowanie, niezależna weryfikacja i pomiary zgodności międzyanotatorskiej zwiększają czas jednostkowy i koszt porcji danych, choć są niezbędne do ograniczania błędów systematycznych. Bez tych praktyk koszt ujawnia się później - jako rework modeli i danych - co zwykle jest jeszcze droższe (Gorgoń i in., 2022).

Nie do pominięcia jest również „podatek techniczny” prywatności. Mechanizmy anonimizacji/maskowania wymagają dodatkowego przetwarzania i często zwiększają objętość plików wyjściowych, co generuje koszty składowania i transferu oraz może wydłużać cały cykl adnotacji; literatura techniczna pokazuje empirycznie, że dobór architektury i parametrów anonimizacji wpływa na rozmiar i przepływ danych, a więc i na ekonomię projektu (Cichowski, Czyżewski, 2013).

Wreszcie, ryzyka integracyjne (vendor lock-in narzędzi, brak kompatybilnych formatów, rozproszona własność danych) powodują, że nawet dobrze zorganizowane procesy etykietowania napotykają bariery przy łączeniu z innymi systemami przedsiębiorstwa. Ujęcia procesowe wskazują, że takie bariery podnoszą koszty transakcyjne między funkcjami/partnerami oraz wydłużają czas dostarczenia efektu, jeśli nie towarzyszy im plan standaryzacji i integracji (Jasińska, 2017).

Podsumowując, główne wady i ograniczenia etykietowania danych wizyjnych układają się w cztery wiązki:

- wysokie CAPEX/OPEX infrastrukturalne i narzędziowe,
- rosnące koszty zgodności i bezpieczeństwa (normy, anonimizacja, audyty),
- długi i niepewny zwrot zależny od stabilności całego łańcucha danych oraz
- niedoszacowane koszty jakości i zmiany (QA, szkolenia, transformacja procesów).

Literatura polskojęzyczna i techniczna wskazuje, że bariery te są częściowo nieusuwalne, ale można je ograniczać poprzez standaryzację formatów i wytycznych, wczesne planowanie zgodności, iteracyjne wdrażanie zmian oraz ścisłe łączenie miar jakości danych z decyzjami finansowymi projektu.

Rozdział 3. Propozycje i warianty optymalizacji procesów zarządzania laboratorium

W niniejszym rozdziale przedstawiono propozycje i warianty stanu docelowego („to-be”) dla laboratorium etykietującego dane wizyjne, oparty na praktykach Agile (Scrum, Kanban, Scrumban) i ukierunkowany na zwiększenie przejrzystości pracy, stabilności przepływu oraz jakości etykiet przy zachowaniu przewidywalności dostaw.

Dobór wariantu przewidziano w oparciu o jawne kryteria: wartość i terminowość, parametry przepływu (lead time, przepustowość, WIP), jakość anotacji, koszty (w tym rework), ryzyka operacyjno-prawne oraz zdolność do skalowania.

3.1. Przegląd zwinnych metod zarządzania w kontekście etykietowania danych

Scrum zapewnia rytm inspekcji i adaptacji (sprinty, przeglądy, retrospektywy) oraz bramkę jakości Definition of Done (Schwaber, Sutherland, 2020). Kanban koncentruje się na wizualizacji pracy, limitach WIP i pomiarze przepływu (lead time, throughput, czas w stanie, wiek zadań) (Kanban University, 2024; Anderson, Carmichael, 2016). Scrumban łączy oba podejścia, umożliwiając utrzymanie cyklicznych przeglądów przy ciągłym „pull” pracy i ograniczaniu WIP, co sprzyja pracy z nierównomiernym napływem zadań (Ladas, 2009; Kniberg, Skarin, 2009). W warunkach zmiennego wolumenu, koniecznych kalibracji oraz rygorów jakościowych rozwiązanie hybrydowe pozwala łączyć elastyczność z przewidywalnością (Reinertsen, 2009).

Dla uporządkowania kluczowych cech metod przygotowano porównanie trzech podejść z perspektywy laboratorium etykietującego. Zwrócono uwagę na rytm planowania, sterowanie obciążeniem, metryki i typowe zastosowania (Schwaber, Sutherland, 2020; Kanban University, 2024).

Tabela 16. Porównanie Scrum / Kanban / Scrumban w laboratorium etykietującym

Kryterium	Scrum	Kanban	Scrumban
Cel nadrzędny	Empiryczne dostarczanie przyrostów w sprintach	Usprawnianie przepływu i przewidywalności	Empiryka + przepływ
Planowanie	Czasowe (sprinty)	Ciągłe (pull)	Sprinty + pull



Kontrola obciążenia	Zakres sprintu	Limity WIP	Zakres + WIP
Metryki wiodące	Cel Sprintu, DoD (ew. velocity)	Lead time, throughput, czas w stanie, wiek zadań	Zestaw mieszany
Zastosowanie	Stały rytm przeglądów, praca porcjowana	Zmienny napływ zadań, skracanie LT	Potrzeba korzyści obu podejść

Źródło: opracowanie własne na podstawie (Schwaber, Sutherland, 2020)

Zestawienie wskazuje, że wybór ramy zarządzania powinien wynikać z natury strumienia pracy i wymaganego poziomu przewidywalności. Scrum porządkuje pracę w porcje czasowe, co sprzyja regularnym kalibracjom instrukcji i komunikacji z interesariuszami, lecz może generować „schodkowość” dostaw (porcje na koniec sprintu) oraz ryzyko przeciążenia, jeżeli zakres sprintu nie uwzględnia zmienności napływu zadań. Kanban działa najlepiej, gdy kluczowe jest skracanie czasu realizacji i redukcja wariacji - limity WIP i ciągły pull równoważą obciążenie wąskich gardeł (szczególnie QA), ale wymagają dojrzałej dyscypliny wizualizacji i codziennej pracy z danymi przepływu. Scrumban łączy zalety obu podejść: zachowuje rytuały inspekcji (planowanie, przeglądy, retrospektywy) i jednocześnie steruje obciążeniem przez WIP oraz pull, co zmniejsza ryzyko przestoju lub „zalania” zespołu.

Na poziomie wdrożeniowym krytyczne jest dopasowanie metryk: w Scrumie warto unikać nadmiernego akcentu na velocity (które bywa niestabilne w pracach usługowych), a w Kanbanie - regularnie przeglądać rozkład lead time’u (percentyle) zamiast średniej. W praktyce laboratoryjnej sprawdzają się hybrydy: rytm kalibracji i przeglądów jakości z Scruma oraz ciągle pobieranie pracy i limity WIP z Kanbanu. (Ladas, 2009; Kniberg, Skarin, 2009).

Poniższa tabela 17 ukazuje podstawowe praktyki Kanban na konkretne czynności w laboratorium, wskazując oczekiwane efekty i sugerowane metryki monitoringu (Anderson, Carmichael, 2016; Kanban University, 2024).



Tabela 17. Kluczowe praktyki Kanban i ich przełożenie na etykietowanie

Praktyka	Zastosowanie w labie	Oczekiwany efekt	Metryka
Wizualizacja przepływu	Kolumny: przygotowanie → anotacja → QA1 → QA2/adjudykacja → publikacja	Przejrzystość, mniej blokad	liczba blokad, czas w stanie
Limity WIP	Limity na kolumnę/osobę	Stabilny czas realizacji	lead time, WIP
Pull	Pobieranie kolejnych pakietów po zakończeniu bieżącego	Mniej rozproszenia	throughput
Jawne zasady	Instrukcje anotacji i polityki QA	Spójność interpretacji	błąd krytyczny %, κ/α
Kadencje	Przegląd przepływu/strategii	Usprawnienia systemowe	trend LT, wykres CFD

Zródło: opracowanie własne na podstawie (Little, 1961, Reinertsen, 2009)

Najsilniejszym regulatorem systemu jest WIP: zmniejszenie limitów na kolumny lub role obniża średni czas realizacji i zmienność wyników przy stałej przepustowości. Kluczowe jest jednak odpowiednie zszywanie wizualizacji z politykami: kolumny powinny odzwierciedlać realne etapy (np. osobno QA1 i adjudykacja), torami mogą być role (anotatorzy, weryfikatorzy) lub klasy usług (np. expedite, fixed-date, standard). Warto zdefiniować zasady wejścia i wyjścia między kolumnami (co minimalnie musi być spełnione), co ogranicza przerwania i cofanie prac.

W warstwie metrycznej, poza lead time'em i throughputem, przydatne jest śledzenie wieku najstarszego zadania (aging), aby aktywnie odytkać zatory, oraz analiza CFD, która ujawnia miejsca narastania WIP. Dla planowania przewidywalnych terminów użyteczne są SLE (Service Level Expectations) oparte na percentylach lead time'u, np. 85% zadań do 5 dni roboczych. Jawne zasady (polityki etykietowania, definicje błędów krytycznych) redukują wariancję interpretacji, a przeglądy przepływu i strategii pozwalają systematycznie usuwać przyczyny blokad, zamiast jedynie reagować na objawy (Anderson, Carmichael, 2016).

Aby ograniczyć rework i zwiększyć spójność, wskazane jest ujednoczenie kryteriów „wejścia” i „wyjścia” pracy. Tabela 18 zawiera minimalne elementy Definition



of Ready i Definition of Done dla laboratorium (PMI, Agile Alliance, 2017; Scrum.org, 2021; Schwaber, Sutherland, 2020).

Tabela 18. DoR i DoD - definicje i minimum treści dla laboratorium

Pojęcie	Definicja	Minimum treści w labie
Definition of Ready (DoR)	Zestaw kryteriów „gotowości” elementu do rozpoczęcia	komplet klas i przykładów, opis edge-cases, zasady anonimizacji, kryteria QA
Definition of Done (DoD)	Zestaw kryteriów „ukończenia” porcji danych	QA zakończone, brak błędów krytycznych, spójność \geq próg κ/α , metadane uzupełnione

Źródło: opracowanie własne na podstawie (PMI, Agile Alliance, 2017; Schwaber, Sutherland, 2020)

DoR i DoD pełnią rolę bramek jakości informacji. DoR chroni zespół przed rozpoczynaniem pracy na niekompletnych wymaganiach: oprócz listy klas i przykładów powinien zawierać przypadki graniczne (edge-cases), decyzje o anonimizacji i minimalnym poziomie jakości wejść (np. rozdzielczość obrazów). Dobrym wzorcem jest checklistowa karta zadania z miejscem na status każdego punktu. DoD ujednocila kryteria gotowości do publikacji: obejmuje wyniki QA (np. brak błędów krytycznych, maksymalny dopuszczalny odsetek błędów istotnych), progi zgodności międzyoceniających (κ/α) i kompletność metadanych (np. wersja instrukcji, osoba weryfikująca). W praktyce bramki powinny być wersjonowane wraz ze zmianami instrukcji, a ich przestrzeganie audytowane na próbkach losowych. Zbyt miękkie DoD prowadzi do wzrostu reworku i eskalacji sporów o interpretację; zbyt twarde - do zatorów, jeśli nie zabezpieczono mocy QA (PMI, Agile Alliance, 2017; Schwaber, Sutherland, 2020).

DoR porządkuje „wejście” pracy i ogranicza przerywanie anotatorów; DoD standaryzuje „wyjście” i skraca akceptację po stronie interesariuszy. W praktyce dobrze zdefiniowane DoR/DoD zmniejszają odsetek poprawek i podnoszą zgodność ocen (Scrum.org, 2021).

Zidentyfikowane problemy zestawiono w tabeli 19 z rekomendowanymi interwencjami oraz miernikami, które pozwalają weryfikować skuteczność działań (Anderson, Carmichael, 2016; Kanban University, 2024).



Tabela 19. Typowe problemy w etykietowaniu i rekomendowane praktyki

Problem	Praktyka	Oczekiwany efekt	Metryka
Długi lead time	Ograniczenie WIP, balans obciążenia	Skrócenie czasu realizacji	lead time, WIP
Niska spójność anotacji	DoR z przykładami, DoD z progami QA	Wyższa zgodność	κ/α , odsetek reworku
„Zatkanie” QA	Pull do QA, klasy usług	Płynniejsze przejścia	czas w stanie QA
Częste zmiany wymagań	Przegląd sprintu/strategii	Wcześniejsze wykrywanie zmian	liczba zmian instrukcji/sprint

Źródło: opracowanie własne na podstawie (Schwaber, Sutherland, 2020)

Długi lead time zwykle wynika nie z pojedynczego etapu, lecz z kumulacji WIP i lokalnych zatorów (np. przeciążone QA). Skuteczna interwencja łączy ograniczenie WIP z równoważeniem obciążenia i dopasowaniem rozmiaru pakietów. Spójność anotacji rośnie wraz z jakością materiałów treningowych (zestaw przykładów, decision log dla sporów) i regularnymi kalibracjami. Zatory w QA zmniejszają się dzięki separacji ról (anotacja vs. weryfikacja), klasom usług i jasnym zasadom priorytetyzacji. Częste zmiany wymagań wymagają jawnej kadencji przeglądów instrukcji oraz ich wersjonowania; minimalizuje to rework przez szybkie zamykanie interpretacji spornej klasy. Rekomendowane jest rozróżnienie mierników wiodących (WIP, wiek zadań, czasy w stanie) i wynikowych (lead time, rework), co ułatwia wczesne wykrywanie problemów.

Redukcja czasu realizacji wymaga w pierwszej kolejności kontroli pracy w toku i wyrównania obciążenia wąskich gardeł (szczególnie QA) (Reinertsen, 2009). Spójność anotacji rośnie, gdy DoR zawiera reprezentatywne przykłady, w tym przypadki graniczne, a DoD definiuje twarde progi jakości (PMI, Agile Alliance, 2017; Schwaber, Sutherland, 2020).

Zastosowanie praktyk Kanban (limity WIP, metryki przepływu, praca „pull”) w połączeniu z rytmem inspekcji Scruma oraz bramkami DoR/DoD tworzy kompletny zestaw narzędzi do stabilizacji procesu etykietowania (Kanban University, 2024; Schwaber, Sutherland, 2020). Dla laboratoriów o zmiennym napływie zadań rekomendowany jest wariant hybrydowy (Scrumban), który pozwala utrzymać regularne przeglądy jakości i elastycznie reagować na zmiany wolumenu, przygotowując grunt pod wybór metody zwinnego zarządzania odpowiedniej dla laboratorium.



3.2. Wybór metody zwinnego zarządzania odpowiedniej dla laboratorium

Wybór docelowego podejścia przeprowadzono jako problem wielokryterialny, aby zrównoważyć wartości niemierzalne i mierzalne (czas, jakość, koszty, ryzyko, skalowalność). Przyjęto paradygmat MCDA i zastosowano procedurę AHP do wyznaczenia wag kryteriów (Trzaskalik, 2014; Roszkowska, 2021; Cabała, 2018; Wachstiel, 2013; Parlińska, 2017). Integrację ryzyka oparto na zasadach zarządzania ryzykiem projektowym (PMI, 2021), a wyniki przedstawiono w tabelach wraz z krótką interpretacją.

3.2.1. Kryteria oceny i ich operacjonalizacja

Kryteria odzwierciedlają tezy: wartość/terminowość, przepływ pracy, jakość etykiet, koszt, ryzyko oraz skalowalność. Definicje i sposób pomiaru zebrano poniżej (Trzaskalik, 2014; Roszkowska, 2021).

Tabela 20. Kryteria decyzyjne i operacjonalizacja

Kryterium	Definicja operacyjna	Miara/źródło danych
Wartość i terminowość	Zdolność do dotrzymania SLA i czasu dostawy porcji danych	lead time (dni), terminowość (% dostaw on-time)
Przepływ pracy	Stabilność i przewidywalność przepływu zadań	przepustowość (porcje/tydz.), wiek zadań, czas w stanie
Jakość etykiet	Spójność i poprawność anotacji	κ/α , odsetek błędów krytycznych (%)
Koszt	Koszt jednostkowy i koszt poprawek	koszt/obraz lub /obiekt; rework (%)
Ryzyko	Prawdopodobieństwo i wpływ zdarzeń niepożądanych	rejestr ryzyk: $P \times I$ (skala 1–5)
Skalowalność	Zdolność do wzrostu wolumenu bez utraty jakości	czas wdrożenia nowego anotatora, stabilność metryk przy +X% wolumenu

Źródło: opracowanie własne na podstawie (Trzaskalik, 2014; Roszkowska, 2021)

Zestaw kryteriów obejmuje zarówno skuteczność dostarczania (SLA), jak i zdrowie systemu pracy (przepływ), jakość produktów (etykiet) oraz ekonomię procesu (koszt). Uwzględnienie ryzyka i skalowalności urealnia decyzję: wariant optymalny dziś



może być nieoptymalny przy wzroście popytu lub pod presją regulacyjną. W pomiarach zaleca się stosowanie percentyli (np. P85 lead time'u) i rozdzielanie mierników wiodących od wynikowych. Należy też zadbać o wiarygodność źródeł: dane operacyjne z narzędzi etykietujących i tablic kanbanowych, audyty QA oraz logi zmian instrukcji. Wrażliwość kryteriów na sezonowość i miks zadań (różne typy anotacji) wymaga normalizacji wyników albo raportowania per klasę zadania (Trzaskalik, 2014; Roszkowska, 2021).

3.2.2. Ważenie kryteriów metodą AHP

Wagi wyznaczono metodą porównań parami (AHP). Macierz porównań zbudowano w oparciu o konsensus zespołu; uzyskano spójność ocen na poziomie $CR < 0,10$ (Cabała, 2018; Roszkowska, 2021).

Tabela 21. Wagi kryteriów (AHP)

Kryterium	Waga
Wartość i terminowość	0,25
Przepływ pracy	0,20
Jakość etykiet	0,20
Koszt	0,15
Ryzyko	0,10
Skalowalność	0,10
Suma	1,00

Źródło: opracowanie własne na podstawie (Cabała, 2018; Trzaskalik, 2014)

Ustawienie najwyższych wag na wartość/terminowość oraz przepływ i jakość odzwierciedlają model usługowy laboratorium: klienci oczekują przewidywalnych dostaw i wysokiej spójności etykiet. W praktyce warto przeprowadzić analizę wrażliwości wag (np. $\pm 20\%$) - jeżeli ranking wariantów pozostaje stabilny, decyzja jest bardziej odporna. Jeżeli nie - należy ponownie zebrać osądy lub rozważyć segmentację (inne wagi dla projektów „regulatorskich” vs. „eksperymentalnych”). $CR < 0,10$ oznacza akceptowalną spójność, ale przy dużych rozbieżnościach między interesariuszami warto zastosować moderowaną rundę konsensusu (Cabała, 2018; Trzaskalik, 2014).



3.2.3. Ocena wariantów i ranking bazowy

Warianty oceniono w skali 1–5 względem każdego kryterium z użyciem danych historycznych i opinii eksperckich.

Tabela 22. Ocena wariantów (1–5) oraz wynik ważony

Wariant	Wartość i terminowość (0,25)	Przepływ (0,20)	Jakość (0,20)	Koszt (0,15)	Ryzyko (0,10)	Skalowalność (0,10)	Suma ważona
Scrum	4	3	4	3	3	3	3,45
Kanban	3	5	3	4	4	4	3,75
Scrumban	4	4	4	4	4	5	4,10

Źródło: opracowanie własne na podstawie (Trzaskalik, 2014; Roszkowska, 2021)

Wynik wskazuje na przewagę wariantu hybrydowego, jednak różnice między Kanbanem a Scrumbanem są mniejsze niż między Scrumem a pozostałymi - sugeruje to, że kluczowa jest kontrola przepływu i WIP. Oceny cząstkowe należy traktować jako przybliżenia: ich wiarygodność zależy od jakości danych wejściowych (np. długości okna obserwacji, jednorodności portfela zadań). Zalecane jest przeprowadzenie scenariuszy alternatywnych (np. większa waga ryzyka w projektach wrażliwych regulacyjnie), aby sprawdzić stabilność rankingu (Trzaskalik, 2014; Roszkowska, 2021).

3.2.4. Rejestr ryzyk i mapa ryzyka

Ryzyka zidentyfikowano w toku analizy procesu. Oszacowano prawdopodobieństwo (P) i wpływ (I) oraz zaplanowano reakcje (PMI, 2021).

Tabela 23. Rejestr kluczowych ryzyk (skala 1–5)

ID	Ryzyko	Opis (przyczyna → skutek)	P	I	Eksp. (P×I)	Reakcja
R1	Niespójność interpretacji	Rozbieżne rozumienie instrukcji → spadek κ/α	3	4	12	DoR z przykładami, kalibracje
R2	Przeciążenie QA	Wąskie gardło w weryfikacji → długi lead time	3	4	12	Limity WIP, klasy usług
R3	Zmiany wymagań	Zmiana klas/wytycznych → rework	4	3	12	Kadencje przeglądów, wersjonowanie



R4	Brak dyscypliny WIP	Nadmierny WIP → opóźnienia i błędy	3	3	9	Polityki WIP, monitorowanie CFD
R5	Rotacja anotatorów	Ubytek kompetencji → spadek jakości	2	4	8	Cross-training, standardy pracy

Źródło: opracowanie własne na podstawie (PMI, 2021)

Najwyższe ekspozycje (12) dotyczą przeciążenia QA i zmian wymagań; obie redukują się dzięki limitom WIP i stałym kadencjom inspekcji (PMI, 2021). Trzy najwyższe ekspozycje mają wspólny mianownik: brak synchronizacji wymagań i przepływu. Rozbieżności interpretacji ograniczają się przez bogate DoR (przykłady, edge-cases) oraz regularne kalibracje i decision log dla sporów. Przeciążenie QA sygnalizuje niewystarczającą separację ról i brak klas usług; rozwiązaniem jest wprowadzenie priorytetów (np. fixed-date dla terminowych zleceń) oraz zarządzanie przekrojowymi kolejkami. Zmiany wymagań będą nieuniknione - dlatego konieczne są jawne kadencje przeglądów instrukcji oraz wersjonowanie, aby minimalizować rozjazdy interpretacji i rework.

Tabela 24. Mapa ryzyka (P×I → kategorie)

Ekspozycja	1–4	5–9	10–25
Kategoria	Niskie	Średnie	Wysokie
Działanie	Monitorować	Plany reakcji	Natychmiastowe działania + właściciel

Źródło: opracowanie własne na podstawie (PMI, 2021)

Ryzyka o ekspozycji ≥ 10 wymagają aktywnego zarządzania i przypisania właścicieli; pozostałe monitoruje się cyklicznie (PMI, 2021).

Matryca P×I jest prostym, lecz skutecznym narzędziem priorytetyzacji. Kategorie powinny mieć przypisane standardy działania: dla „wysokich” - obowiązkowy plan reakcji, właściciel, termin przeglądu i kryterium zamknięcia; dla „średnich” - uzgodniony próg eskalacji; dla „niskich” - rytm monitoringu. W praktyce przydatna jest wizualizacja heatmapy i cotygodniowy przegląd stanu ryzyk w ramach przeglądu przepływu. Ważne jest rozpoznanie współzależności: ryzyka QA i zmian wymagań często się nasilają wzajemnie, więc działania powinny być skoordynowane.

3.2.5. Korekta wyniku o ekspozycję na ryzyko

Dla uwzględnienia wrażliwości na ryzyko zastosowano korektę $Wynik_{adj} = Wynik_{baz} \times (1 - \lambda \cdot r)$, gdzie r to znormalizowana ekspozycja (średnia $P \times I / 25$), a $\lambda = 0,2$.

Tabela 25. Ranking po korekcie ryzykiem

Wariant	Średnia ekspozycja ($P \times I$)	Wsp. korekty $(1 - \lambda \cdot r)$	Wynik skorygowany
Scrum	12	0,904	3,12
Kanban	10	0,920	3,45
Scrumban	9	0,928	3,80

Źródło: opracowanie własne na podstawie (PMI, 2021; Roszkowska, 2021)

Po uwzględnieniu ekspozycji na ryzyko utrzymuje się przewaga Scrumban nad Kanbanem; Scrum pozostaje trzeci z uwagi na wrażliwość na zmiany wymagań w oknach sprintów (PMI, 2021; Roszkowska, 2021). Wariant hybrydowy lepiej radzi sobie z ekspozycją na kluczowe ryzyka (QA, zmiany wymagań). Współczynnik λ powinien być uzgadniany z interesariuszami - wyższa awersja do ryzyka (np. w projektach regulowanych) zwiększy „karę” za ekspozycję i może przestawić ranking. Dla pełniejszego obrazu warto wykonać analizę wrażliwości na λ oraz scenariusze alternatywne; tam, gdzie dostępne są dane historyczne, przydatna jest symulacja Monte Carlo na rozkładach lead time’u i obciążenia QA.

Zastosowanie MCDA (AHP) umożliwiło przejrzyste porównanie wariantów „to-be” i wybór rozwiązania równoważącego wartość, przepływ, jakość, koszty, ryzyko i skalowalność. Ranking bazowy i skorygowany ryzykiem wskazuje na przewagę wariantu hybrydowego Scrumban. Rekomenduje się cykliczną weryfikację wag i ocen oraz utrzymywanie rejestru ryzyk jako żywego dokumentu wraz z przeglądami przepływu. (Trzaskalik, 2014; Roszkowska, 2021; PMI, 2021).

3.3. Optymalizacja przepływu pracy i redukcja kosztów

W podrozdziale zaprojektowano stan docelowy (to-be) strumienia wartości laboratorium etykietującego dane wizyjne oraz wskazano działania prowadzące do skrócenia czasu realizacji, podniesienia przewidywalności i obniżenia kosztu jednostkowego. Postępowanie oparto na mapowaniu strumienia wartości (VSM),



wyznaczeniu czasu taktu i reguł balansowania obciążenia, a także na zastosowaniu automatyzacji (pre-labeling, active learning, słaby nadzór) skorelowanych z kontrolą jakości (Rother, Shook, 2019; Womack, Jones, 2012; Reinertsen, 2009; Settles, 2009; Ratner et al., 2017; Northcutt et al., 2021).

3.3.1. VSM to-be – cele przepływu i parametry strumienia

W pierwszym kroku opracowano mapę to-be z etapami: przygotowanie danych → anotacja → QA1 → adjudycacja/QA2 → publikacja. Dla każdego etapu zdefiniowano cele, limity WIP oraz wskaźniki wyniku.

Tabela 26. VSM to-be – etapy, cele i wskaźniki przepływu

Etap	Cel to-be	Docelowe parametry przepływu	Wskaźniki wyniku
Przygotowanie	Standaryzacja paczek (rozmiar, klasy, przykłady, wersja instrukcji)	WIP \leq 1 dzień taktu; czas w stanie \leq 1 dzień	% paczek on-time
Anotacja	Praca pull w paczkach stałej wielkości	WIP/anotator \leq 2 paczki; brak multitasking	lead time paczki; throughput (paczek/tydz.)
QA1	Kontrola na próbie + reguły błędów krytycznych	WIP \leq ½ dnia taktu; AQL dostosowany do ryzyka	błędy krytyczne (%)
Adjudycacja/QA2	Rozstrzygnięcie rozbieżności i kalibracja	WIP \leq ½ dnia taktu; decyzja time-boxed	zgodność κ/α
Publikacja	Eksport + pełne metadane i wersjonowanie	czas w stanie \leq 1 dzień; zero „odbić” do QA	terminowość; rework (%)

Zródło: opracowanie własne na podstawie (Rother, Shook, 2019)

Mapa to-be zakłada stabilizację przepływu poprzez ograniczenie WIP i standaryzację wejść/wyjść (DoR/DoD). Największe skrócenia czasu uzyskuje się zwykle dzięki redukcji oczekiwań międzyetapowych, a nie przez samo przyspieszanie



anotacji (Rother, Shook, 2019). Zgodnie z zasadami przepływu, mniejszy WIP obniża wariację lead time'u i poprawia przewidywalność dostaw (Reinertsen, 2009).

3.3.2. Czas taktu i balansowanie obciążenia

Czas taktu przyjęto jako metronom przepływu: takt = dostępny czas produkcyjny/popyt. Dla środowiska usługowego rekomenduje się takt dzienny/tygodniowy z korektą o okna niedostępności (kalibracje, przerwy techniczne) oraz stały pitch (rytualny „zrzut” paczek).

Tabela 27. Takt i zasady balansowania

Element	Założenie to-be	Zastosowanie
Dostępny czas	Np. 6 h netto/dzień/osobę	Bez spotkań i przerw
Popyt	Np. 1200 obrazów/tydz. → 240/dzień	Przeliczenie na paczki stałej wielkości
Takt	$6 \text{ h} \times \text{liczba os\o{b}} / 240$	Wyznaczenie rytmu paczek
Balans	line balancing względem obciążeń	Przesunięcia ról, wyrównanie paczek
Pitch	Stały „zrzut” (np. co 2 h)	Synchronizacja anotacji z QA

Źródło: opracowanie własne na podstawie (Womack, Jones, 2012)

Takt jest celem operacyjnym, niesztynym limitem. Kluczowe jest powiązanie wielkości paczek z pitchem, aby nie „zalewać” QA porcjami pracy. Niedopasowanie taktu ujawnia się wzrostem wieku zadań i rozjechaniem percentyli lead time'u (Womack, Jones, 2012).

3.3.3. Automatyzacje i pre-labeling

Zaprojektowano trzy komplementarne dźwignie: pre-labeling (auto-adnotacje do korekty), active learning (priorytetyzacja próbek) i słaby nadzór (programowalne etykiety/reguły) oraz uzupełniająco czyszczenie etykiet metodami confident learning.

Tabela 28. Automatyzacje – zastosowania, wymagania, efekty, ryzyka

Technika	Zastosowanie	Wymagania	Oczekiwany efekt	Ryzyka
Pre-labeling (np. w CVAT/Label Studio)	Automatyczne maski/ramki do korekty	Model bazowy, integracja backend ML	Mniej klików; krótszy czas paczki	Uprzedzenia modelu; dryf klas (CVAT Team, 2024; Label Studio, 2024)
Active learning (niepewność, core-set)	Wybór najbardziej informacyjnych próbek	Pętla trening–predykcja–selekcja	Mniej próbek do podobnej jakości modelu	Bias selekcyjny; złożoność (Settles, 2009; Sener, Savarese, 2018)
Słaby nadzór (Snorkel)	Reguły/heurystyki generujące etykiety szkicowe	Repo reguł + walidacja	Szybsze pozyskanie etykiet	Błędy systematyczne; potrzeba adjudykacji (Ratner et al., 2017)
Confident learning (czyszczenie)	Identyfikacja błędnych etykiet	Metryki pewności i pipeline QC	Spadek reworku w QA	Falszywe pozytywy „błędów” (Northcutt et al., 2021)

Źródło: opracowanie własne na podstawie (Sener, Savarese, 2018)

Pre-labeling daje najszybszą poprawę produktywności przy detekcji/segmentacji, o ile modele są adekwatne do domeny. Active learning ogranicza wolumen ręcznej anotacji przy zachowaniu jakości modelu; rekomenduje się rozpoczynać od strategii niepewności i rozszerzać o metody dywersyfikujące (Sener, Savarese, 2018). Słaby nadzór sprawdza się tam, gdzie wiedza ekspercka daje się zapisać w regułach, jednak wymaga ścisłej współpracy z QA, aby ograniczać błędy systematyczne. Czyszczenie etykiet redukuje ukryty koszt błędów, który narasta w dole strumienia (Northcutt et al., 2021).



3.3.4. Ekonomia usprawnień – koszt i zwrot

Wdrożenia podzielono na CAPEX (integracje, infrastruktura) i OPEX (szkolenia, inference, kalibracje). Skutki rozpatrzono w horyzoncie krótkim i dłuższym.

Tabela 29. Struktura kosztów i spodziewane efekty

Pozycja	Charakter	Przykłady	Efekt krótki	Efekt długi
Narzędzia/integracje	CAPEX/OPEX	Serwer inference, wtyczki ML	Krótszy czas anotacji	Spadek kosztu jednostkowego
Szkolenia/kalibracje	OPEX	DoR/DoD, checklista QA	Mniej reworku	Wyższa zgodność κ/α
Pre-labeling	OPEX	Batch inference	Szybsze paczki	Mniej pracy ręcznej
Active learning	OPEX	Pętle selekcji	Mniej próbek	Krótsze iteracje R&D
Słaby nadzór	CAPEX/OPEX	Repo reguł	Szybkie etykiety szkieletowe	Skalowanie bez liniowego wzrostu etatów

Źródło: opracowanie własne na podstawie (Sener, Savarese, 2018)

Integracje i automatyzacje są kosztowniejsze na starcie, lecz przynoszą nieliniowe korzyści skali. Kalibracje i standardy działają jako prewencja reworku, co natychmiast odciąża QA i skraca lead time.

3.3.5. Harmonogram wdrożenia to-be (0–90 dni)

Plan wdrożenia podzielono na trzy fale z miernikami sukcesu.

Tabela 30. Roadmap wdrożenia i wskaźniki

Fala	Zakres	Dostarczenia	Mierniki (baseline → cel)
0–30 dni	VSM to-be, DoR/DoD 1.0, tablice Kanban	Mapa stanu docelowego, polityki WIP, SLE	Lead time P85: X→X-20%; WIP: X→X-30%



31–60 dni	Pre-labeling + QA sampling	Inference batch, AQL dla QA	Czas anotacji/paczka: X→X–25%; błędy kryt.: X→X–50%
61–90 dni	Active learning + Snorkel pilotaż	Pętla selekcji, 3–5 LF	Wolumen ręcznej anotacji: X→X–30%; κ/α : +0,05

Źródło: opracowanie własne na podstawie (CVAT, 2024)

Kolejność minimalizuje ryzyko: najpierw stabilizacja przepływu i definicji (VSM, DoR/DoD), następnie automatyzacje o bezpośrednim wpływie na koszt (pre-labeling), a na końcu techniki bardziej złożone (active learning, słaby nadzór). Mierniki oparte na procentylach lepiej oddają doświadczenie klienta niż średnie (CVAT, 2024).

Optymalizacja przepływu w oparciu o VSM, takt i limity WIP, uzupełniona automatyzacjami (pre-labeling, active learning, słaby nadzór), pozwala skrócić lead time, obniżyć koszt jednostkowy i zwiększyć spójność etykiet przy rosnącej skali. Zapropionowany harmonogram umożliwi wdrożenie w ciągu 90 dni przy jednoczesnym mierzeniu efektów.

3.4. Poprawa jakości etykietowania poprzez kontrolę i podwójną weryfikację

W podrozdziale ujęto trzy uzupełniające się filary jakości:

- kontrolę próbkową partii według atrybutów (AQL) na podstawie ISO 2859-1,
- podwójną weryfikację z adjudykacją jako mechanizm redukcji błędów i pozyskiwania wiedzy o niejednoznacznościach,
- cykliczne kalibracje wspierane miarami zgodności międzyoceniających (Cohen's κ ; Krippendorff's α) i metrykami specyficznymi dla obrazów (np. Dice/IoU przy segmentacji).

Uzupełniająco wykorzystano praktyki kontroli jakości z crowdsourcingu: złote zadania, ukryte testy, ważenie annotatorów i adaptacyjne próbkowanie (Daniel et al., 2018; Oleson et al., 2011; Sheshadri, Lease, 2013; Aroyo, Welty, 2014; Davani et al., 2022; Yang et al., 2023).



3.4.1. Kontrola próbkowa partii (AQL, ISO 2859-1)

Kontrola na próbie według atrybutów pozwala akceptować lub odrzucać partie etykiet na podstawie z góry ustalonego poziomu jakości akceptowalnej (AQL) i planu pobierania próby (ISO 2859-1:1999). W laboratorium etykietującym dane wizyjne „partią” jest paczka adnotacji z jednorodnego okresu/klasy zadań.

Tabela 31. Przykładowa polityka AQL dla paczek etykiet (wg ISO 2859-1:1999, adaptacja)

Klasa ryzyka błędu	Przykłady błędów	Poziom kontroli	AQL (docelowe)	Plan próbkowania (przykład)	Reguła akceptacji
Krytyczna	Naruszenie prywatności; błędna klasa w danych regulowanych	Srogi (inspection level II–III)	0,1–0,4%	n=125; c=0	Akceptacja wyłącznie przy 0 błędach
Poważna	Pomyłka klasy obiektu; rażąca niekompletność maski	Zwykły (level II)	1,0%	n=80; c=1	Odrzucenie przy ≥ 2 błędach
Drobna	Drobna granica maski; nieistotne metadane	Zmniejszony (level I)	2,5–4,0%	n=50; c=3	Odrzucenie przy ≥ 4 błędach

Źródło: opracowanie własne na podstawie (ISO 2859-1:1999)

ISO 2859-1 definiuje rodzinę planów próbkowania; wartości n (wielkość próby) i c (liczba dopuszczalnych niezgodności) dobiera się do rozmiaru partii i przyjętego AQL. W praktyce laboratorium zaleca się różnicowanie poziomów kontroli według wagi konsekwencji błędu: klasy krytyczne kontroluje się częściej i surowiej (c=0), a klasy drobne - rzadziej i z wyższym progiem akceptacji. Taki podział zmniejsza koszt QA przy zachowaniu ryzyka na akceptowalnym poziomie (ISO 2859-1:1999).

3.4.2. Podwójna weryfikacja i adjudykacja

Podwójna, niezależna weryfikacja (double-blind) pozwala oszacować zgodność międzyoceniającą oraz wychwycić niejednoznaczności, które nie ujawniają się przy

pojedynczym przeglądzie. Sporne przypadki rozstrzyga się w ścieżce adjudykacji przez doświadczonego sędziego/komitet, dokumentując decyzję w rejestrze (decision log), co zasila późniejsze kalibracje.

Tabela 32. Warianty kontroli jakości a koszt i efekt

Wariant	Opis	Zastosowanie	Zalety	Ograniczenia
QA 1- stopniowe (sampling)	Przegląd próbki wg AQL	Gdy ryzyko jest niskie/średnie	Niski koszt, szybka informacja	Szum w estymacji, mniejsza czułość na rzadkie błędy
QA 2- stopniowe (double- blind)	Dwie niezależne adnotacje + porównanie	Klasy ryzyka średnie/wysokie	Pomiar zgodności, lepsze wychwytywanie sporów	Wyższy koszt godzinowy
Adjudykacja	Rozstrzygnięcie niezgodnych przypadków	Gdy brak konsensusu	Standaryzacja i nauka z błędów	Wydłużenie cyklu bez dyscypliny czasowej
QA adaptacyjne	Zmienny próg próbki wg jakości	Portfele o zmiennej jakości	Skierowanie wysiłku na obszary ryzyka	Złożoność konfiguracji i monitoringu

Źródło: opracowanie własne na podstawie (Oleson et al., 2011; Daniel et al., 2018)

Dobór wariantu należy uzależnić od wagi konsekwencji błędu i obserwowanej zmienności jakości. Wysoka niezgodność między annotatorami jest sygnałem do:

- podniesienia intensywności kontroli,
- przeglądu instrukcji i przykładów,
- uruchomienia kalibracji.

Praktyki z crowdsourcingu (złote zadania, testy ukryte, targetowane próbkowanie) pozwalają obniżyć koszt QA2, kierując podwójną weryfikację tam, gdzie ryzyko jest największe (Oleson et al., 2011; Daniel et al., 2018).

3.4.3. Miary zgodności międzyoceniających i metryki dla obrazów

Dla zadań nominalnych (np. klasy obiektu) stosowany jest współczynnik Cohen's κ , który koryguje zgodność o przypadkowe dopasowania (Cohen, 1960). Dla układów z więcej niż dwiema osobami oceniającymi i różnymi skalami (nominalna, porządkowa, metryczna) zalecany jest Krippendorff's α , odporny na brakujące dane (Krippendorff, 2013/2022). W segmentacji obrazów pomiar zgodności uzupełnia się o Dice/IoU i – w razie potrzeby - metryki brzegowe (Yang et al., 2023).

Tabela 33. Dobór miary zgodności do typu zadania

Typ zadania	Miara	Kiedy stosować	Uwagi metodyczne
Klasyfikacja nominalna (2 oceniających)	Cohen's κ	Dwie niezależne adnotacje	Wrażliwy na rozkład klas; raportować także zgodność „surową”
Klasyfikacja nominalna (≥ 3 oceniających)	Krippendorff's α (nominal)	Więcej niż dwóch oceniających, brakujące dane	Obsługuje przekodowania i brak obserwacji
Skale porządkowe/ciągłe	Krippendorff's α (ordinal/interval)	Cięcia prógowe, oceny stopniowane	Wymaga dobrania funkcji odległości
Segmentacja/detekcja	IoU/Dice + α lub κ na etykietach instancji	Maski/ramki obiektów	Rozważyć metryki brzegowe i tolerancję odległości

Źródło: opracowanie własne na podstawie (Cohen, 1960; Krippendorff, 2013/2022; Yang et al., 2023)

W praktyce wskazane jest raportowanie pary: „zgodność międzyoceniających” (κ/α) oraz jakości geometrycznej (IoU/Dice) dla zadań przestrzennych. Dla problemów z rzadkimi klasami κ może zaniżyć ocenę - wtedy przydatne są dodatkowe przekroje wyników (per klasa) oraz analiza niepewności adnotatorów (Cohen, 1960; Krippendorff, 2013/2022; Yang et al., 2023).

Tabela 34. Przykładowe progi decyzyjne i zastosowanie w procesie

Miara	Próg roboczy (orientacyjny)	Zastosowanie w bramkach
Krippendorff's α	$\geq 0,80$ ($\geq 0,667$ – warunkowo)	Akceptacja definicji klas/wytycznych; poniżej - obowiązkowa kalibracja
Cohen's κ	Ustala się projektowo (np. $\geq 0,70$)	Wejście do stanu „Done” dla paczki po QA2
IoU/Dice	Ustala się per projekt/klasa	Kryteria QA dla segmentacji/detekcji

Źródło: opracowanie własne na podstawie (Yang et al., 2023, Krippendorff, 2013/2022)

Wartości progowe powinny być ustalane projektowo i potwierdzane empirycznie. Rekomendacja $\alpha \geq 0,80$ jako standardu jakości oraz $\alpha \geq 0,667$ „warunkowo” bywa przyjmowana w analizie treści (Krippendorff, 2013/2022). W zadaniach obrazowych progi IoU/Dice zależą od tolerancji błędu i przeznaczenia modelu; często sensowne jest różnicowanie progów dla obiektów małych vs. dużych (Yang et al., 2023).

3.4.4. Jakość w środowisku zespołowym/crowd: złote zadania, adaptacja próby, konsensus

Kontrola jakości w środowiskach wieloosobowych korzysta z pakietu narzędzi: złote zadania (zadania z odpowiedzią referencyjną), testy ukryte, ważenie annotatorów według trafności oraz adaptacyjne próbkowanie kierowane wskaźnikami ryzyka (Daniel et al., 2018; Oleson et al., 2011). Ustalenie konsensusu nie powinno być redukowane do prostej większości; dla zadań subiektywnych warto zachować rozkład opinii (Davani et al., 2022) i traktować niejednoznaczność jako sygnał, a nie błąd (Aroyo, Welty, 2014). Ramy benchmarkowe (np. SQUARE) umożliwiają porównywalny pomiar jakościową „siłą” konsensusu (Sheshadri, Lease, 2013).

Tabela 35. Mechanizmy QA w środowisku wieloosobowym

Mechanizm	Cel	Jak wdrożyć	Uwagi praktyczne
Złote zadania / testy ukryte	Walidacja kompetencji	Wtrącenie niewielkiego odsetka zadań referencyjnych	Aktualizować „złoto” po zmianach instrukcji



Ważenie annotatorów	Zmniejszanie wpływu słabych etykiet	Wagi ~ trafność na złotych zadaniach	Chronić przed błędnym karaniem ekspertów od rzadkich klas
Adaptacyjne próbkowanie	Skierowanie QA tam, gdzie ryzyko	Większa próba dla anotatorów/klas o słabszych wynikach	Bilansować kosztem i opóźnieniami
Model konsensusu	Lepsze scalanie opinii	Większość ważona, modele probabilistyczne	Zachować rozkłady dla zadań subiektywnych

Źródło: opracowanie własne na podstawie (Aroyo, Welty, 2014; Davani et al., 2022)

Mechanizmy te obniżają koszt, przesuując uwagę na miejsca największego ryzyka. Zachowanie informacji o rozbieżności (np. dystrybucji etykiet) pomaga odróżnić błędy od prawdziwej wieloznaczności danych (Aroyo, Welty, 2014; Davani et al., 2022).

3.4.5. Kalibracje: rytm, artefakty i zamknięcie pętli

Kalibracje są cyklicznymi spotkaniami techniczno-merytorycznymi służącymi „domykaniu” rozbieżności i aktualizacji materiałów referencyjnych. Wskazane jest utrzymywanie logu decyzji, repozytorium przykładów wzorcowych/edge-cases i wersjonowania instrukcji.

Tabela 36. Minimalny standard kalibracji

Element	Co obejmuje	Dowód/artefakt	Wpływ na proces
Rytm	Np. co tydzień (lub po przekroczeniu progu sporów)	Harmonogram i protokół	Stale obniżanie sporów i reworku
Zakres	Sporne klasy, najczęstsze błędy, nowe przykłady	Zaktualizowana instrukcja, zestaw przykładów	Wyższa zgodność κ/α ; mniej błędów krytycznych
Decyzje	Ustalenia brzegów klas, zasady rozstrzygania	Decision log (z datą i właścicielem)	Audytowalność zmian
Domknięcie	Wdrożenie zmian i komunikacja na tablicy	Release notes do instrukcji	Jasność kryteriów DoR/DoD

Źródło: opracowanie własne



Kalibracje powinny być sprzężone z metrykami: spadek odsetka sporów, wzrost κ/α , skrócenie czasu adjudykacji. Dokumentowanie decyzji zmniejsza dryf interpretacji i przyspiesza onboardowanie kolejnych osób.

Spójne połączenie kontroli próbkowej AQL (ISO 2859-1), podwójnej weryfikacji z adjudykacją oraz systematycznych kalibracji pozwala istotnie obniżyć ryzyko błędów krytycznych i koszt poprawek, a także stabilizuje zgodność międzyoceniających. Uzupełnienie o praktyki crowdsourcingowe (złote zadania, adaptacja próby, modele konsensusu) umożliwia skalowanie jakości przy akceptowalnym koszcie.

W części poświęconej zarządzaniu laboratorium etykietującym dane wizyjne zaprojektowano spójny model operacyjny łączący podejścia zwinne, inżynierię przepływu i statystyczną kontrolę jakości. Dla strumienia pracy o zmiennym napływie zadań wskazano wariant hybrydowy Scrumban (rytm inspekcji i adaptacji + wizualizacja pracy, limity WIP, metryki przepływu), ustandaryzowano bramki DoR/DoD oraz określono mierniki wiodące (wiek zadań, czasy w stanie, percentyle lead time'u).

Dobór metody wykonano jako problem MCDA z ważeniem AHP, definiując kryteria: wartość/terminowość, przepływ, jakość, koszt, ryzyko, skalowalność. Ranking bazowy i po korekcie ryzykiem potwierdził przewagę Scrumban, a rejestr i mapa ryzyk ukierunkowały działania zapobiegawcze (przeciążenie QA, zmiany wymagań, dyscyplina WIP).

Zaprojektowano stan docelowy VSM to-be z czasem taktu, zasadami balansowania obciążenia i limitami WIP, a także pakiet automatyzacji: pre-labeling, active learning, słaby nadzór oraz czyszczenie etykiet (confident learning). Przedstawiono roadmapę 0–90 dni i metryki sukcesu (lead time P85, WIP, rework, κ/α).

System jakości oparto na AQL (ISO 2859-1), podwójnej weryfikacji z adjudykacją oraz cyklicznych kalibracjach, z miarami zgodności (Cohen's κ , Krippendorff's α) i metrykami przestrzennymi (IoU/Dice). Oczekiwany rezultat to krótszy i bardziej przewidywalny lead time, niższy koszt jednostkowy oraz wyższa spójność etykiet, co przekłada się na wzmocnienie konkurencyjności jednostki laboratorium.



Rozdział 4. Wdrożenie wybranych metod zarządzania w laboratorium

Niniejszy rozdział przedstawia założenia i realizację studium przypadku firmą inżynierską a uczelnią techniczną oraz wyniki oceny efektów wdrożenia usprawnień procesów etykietowania danych. Zbadano zastosowanie zwinnych metod zarządzania w optymalizacji procesu etykietowania danych wizyjnych na potrzeby uczenia maszynowego w warunkach rzeczywistych.

Celem badania było określenie wielowymiarowego wpływu usprawnień na funkcjonowanie jednostki laboratoryjnej - w wymiarze operacyjnym, jakościowym, kosztowym i serwisowym. Punkt wyjścia stanowi hipoteza, że spójny pakiet działań (Scrumban, limity WIP, DoR/DoD, VSM to-be, AQL, podwójna weryfikacja z adjudykacją, pre-labeling) skraca czas realizacji, stabilizuje przepływ i podnosi zgodność międzyoceniających, obniżając równocześnie koszt jednostkowy.

Laboratorium stanowi wspólne przedsięwzięcie firmy inżynierskiej i uczelni technicznej, działające w modelu partnerskim. Firma inżynierska dostarcza dane wejściowe, sprzęt i licencje, zapewnia dostęp do zasobów przez VPN oraz finansuje i nadzoruje dwóch koordynatorów merytorycznych. Uczelnia techniczna udostępnia przestrzeń biurową (ok. 160 m²/45 stanowisk), infrastrukturę sieciową i obsługę administracyjno-kadrową, zatrudniając studentów-anotatorów i weryfikatorów (status studenta, wiek do 26 lat).

Przedmiotem badań jest kompletny strumień wartości etykietowania: przygotowanie paczek danych, anotacja, kontrola jakości jedno- i dwustopniowa (QA1, QA2 z adjudykacją), kalibracje oraz publikacja wyników wraz z metadanymi. Analizie poddaje się zarówno sposób organizacji pracy (tablice z limitami WIP i przepływem pull), jak i bramki jakości (DoR/DoD, AQL) oraz wybrane automatyzacje (pre-labeling do korekty).

Główne pytanie badawcze brzmi: *jaki wpływ na funkcjonowanie jednostki laboratorium ma usprawnienie procesów etykietowania danych?* Oczekuje się odpowiedzi w postaci mierzalnych zmian wskaźników przepływu, jakości i ekonomii, a także w postaci efektów serwisowych widocznych w realizacji SLA.

Zakres badania obejmuje porównanie stanu bazowego i stanu po wdrożeniu na porównywalnych oknach czasowych. Zdefiniowano zestaw metryk: lead time (P50/P85), cycle time i czasy w stanie, throughput (paczek/tydz.), WIP i aging; jakość (błędy



krytyczne, rework, κ/α ; dla zadań przestrzennych także IoU/Dice); serwis (SLA on-time); ekonomia (koszt/h, koszt/etykietę z dekompozycją na robociznę i koszty ogólne).

Pozyskanie danych odbyło się wyłącznie w bezpiecznym środowisku firma inżynierska (dostęp przez VPN), z wykorzystaniem dzienników narzędzi anotacyjnych, tablic przepływu, rejestrów QA i logów decyzji z kalibracji. Zastosowano triangulację: ujęcia ilościowe (metryki, testy różnic) uzupełniono ujęciami jakościowymi (wnioski z przeglądów przepływu i sesji kalibracyjnych), przy zachowaniu ograniczeń wynikających z poufności danych.

Wdrożenie zrealizowano w trzech etapach (0–90 dni). W pierwszej ustalono baseline metryk, wykonano VSM to-be, uruchomiono tablice z limitami WIP, zdefiniowano DoR/DoD i cykl przeglądów przepływu/jakości. W drugiej wprowadzono kontrolę próbkową AQL (QA1), podwójną weryfikację z adjudycją (QA2), wersjonowanie instrukcji i decision log; rozpoczęto pre-labeling. W trzeciej ustalono SLE (percentylowe oczekiwania czasu dostawy), ustabilizowano kadencję kalibracji i zoptymalizowano wielkość paczek oraz pitch zasilania QA. Całość prowadzono przy ścisłej segregacji ról (anotacja, QA, adjudycją) oraz kontroli dostępu do danych (VPN, konta domenowe).

Zastosowano:

- tablicę pracy z kolumnami przygotowanie → anotacja → QA1 → adjudycją/QA2 → publikacja oraz limitami WIP per kolumna/rola,
- platformę anotacyjną z pre-labelingiem (auto-maski/ramki do korekty) i integracją batch-inference,
- repozytorium instrukcji i przykładów (w tym edge-cases), rejestr sporów i decision log,
- zestaw kontroli jakości: plany AQL, złote zadania/testy ukryte, mierniki κ/α i – dla segmentacji – IoU/Dice,
- zaplecze operacyjne: konta domenowe, VPN, stacje robocze i licencje dostarczone przez firmę inżynierską, biuro i obsługę kadrową po stronie uczelni technicznej.

Model to-be odwzorowano w notacji BPMN z czterema bramkami:

- Gate-0 (DoR) - kompletność i wersja instrukcji,



- Gate-1 (QA1/AQL) - kontrola próbkowa i reguły błędów krytycznych;
- Gate-2 (QA2/adjudykacja) - rozstrzygnięcie rozbieżności i pomiar zgodności (κ/α ; IoU/Dice),
- Gate-3 (DoD/publikacja) - końcowy przegląd jakości i metadanych.

Przepływ sterowany był pull i limitami WIP; wielkość paczek i pitch dobrano do taktu wynikającego z popytu. Monitorowane są metryki wiodące (WIP, wiek zadań, czasy w stanie, CFD) i wynikowe (lead/cycle time - w percentylach, rework, SLA).

Oceny dokonano w układzie „przed-po” na porównywalnych oknach czasowych. Oczekiwane i uzyskane zmiany obejmują: spadek lead time'u P85 o ~20–30%, wzrost throughputu o ~15–25%, redukcję reworku o ~40–60%, wzrost κ/α o ~0,05–0,10 oraz spadek odsetka błędów krytycznych o $\geq 50\%$. Pre-labeling skrócił czas anotacji paczki (typowo o 20–25%) bez pogorszenia progów IoU/Dice i zgodności; ustabilizowano SLA on-time (wzrost o ≥ 10 p.p.). W kalkulacji kosztu/etykiety wykazano spadek wynikający z krótszego czasu jednostkowego i mniejszego udziału poprawek, przy stałym nadzorze bezpieczeństwa danych (VPN).

Stabilny przepływ i transparentny system jakości (AQL + QA2 + kalibracje) podnoszą wiarygodność wobec klientów oraz ułatwiają negocjowanie SLA. Przewidywalny lead time i niższy koszt/etykiety zwiększają konkurencyjność cenową, a możliwość skalowania (pipeline studentów, pre-labeling) ułatwia absorpcję zmiennego popytu. Uporządkowane artefakty (wersjonowane instrukcje, decision log) skracają onboardowanie i obniżają ryzyko operacyjne, wzmacniając pozycję firmy inżynierskiej i uczelni technicznej jako partnera zdolnego do terminowej realizacji etykietowania danych wizyjnych na potrzeby ML.

4.1. Opis procesu wdrożenia

Wdrożenie przeprowadzono w laboratorium firmy inżynierskiej i uczelni technicznej, w którym firma inżynierska zapewnia dane, narzędzia, infrastrukturę dostępową (VPN), sprzęt i koordynację merytoryczną, a uczelnia techniczna udostępnia przestrzeń biurową, łączność oraz obsługę kadrowo-administracyjną studentów-anotatorów. Celem było usystematyzowanie przepływu, podniesienie jakości i przewidywalności dostaw oraz obniżenie kosztu jednostkowego etykiety. Działania zogniskowano wokół Scrumban (wizualizacja pracy, limity WIP, rytuały), bramek jakości DoR/DoD i planów AQL, ścieżki podwójnej weryfikacji z adjudykacją oraz cyklu kalibracji. W tabeli 37 przedstawiono zakres zmian w czterech osiach: przepływ, jakość,



automatyzacje i bezpieczeństwo. Każdy element powiązano z artefaktami i miernikami akceptacji.

Tabela 37. Zakres wdrożenia w laboratorium firmy inżynierskiej i uczelni technicznej (obszar → zmiana → artefakty → mierniki)

Obszar	Zmiana docelowa (to-be)	Artefakty/standardy	Mierniki akceptacji
Przepływ	Praca pull, limity WIP na etapy/role, stałe paczki	Tablica: przygotowanie → anotacja → QA1 → adjudykacja/QA2 → publikacja; polityki WIP	LT P85 ↓, WIP/aging ↓, throughput ↑
Jakość	Bramki DoR/DoD, AQL, QA2 + adjudykacja, kalibracje	DoR/DoD 1.0; plany AQL; decision log; biblioteka przykładów/edge-cases	Błędy krytyczne ↓, rework ↓, κ/α ↑, IoU/Dice \geq próg
Automatyzacje	Pre-labeling do korekty na stabilnych klasach	Pipeline batch-inference; raport IoU/Dice na pre-labelu	Czas anotacji/paczka ↓ przy stabilnym κ/α
Bezpieczeństwo	Praca wyłącznie przez VPN i konta domenowe	Rejestry dostępu; polityki anonimizacji	Zero incydentów dostępu; pełna zgodność procedur

Zródło: opracowanie własne

Zakres scala zarządzanie przepływem z kontrolą jakości i bezpieczeństwem danych. W praktyce największą dźwignię operacyjną zapewnia ograniczenie WIP i standaryzacja wejść/wyjść (DoR/DoD), natomiast AQL i QA2 działają jako filtr kosztownych błędów, zanim trafią do publikacji. Pre-labeling wprowadzono dopiero po ustabilizowaniu jakości, aby automatyzacja nie wzmacniała błędów systematycznych. Wszystkie zmiany wiążą się z miernikami akceptacji, co umożliwia rozliczanie efektu „przed–po”.

Harmonogram podzielono na trzy fale 0–90 dni. Dla każdej fali zdefiniowano dostarczenia i kryteria akceptacji wprost w wskaźnikach. Dane przedstawiono w tabeli 38.



Tabela 38. Harmonogram wdrożenia 0–90 dni (dostarczenia i kryteria akceptacji)

Fala	Okres	Główne dostarczenia	Kryteria akceptacji (KPI)
I	0–30 dni	Tablica z WIP; DoR/DoD 1.0; VSM to-be; baseline metryk	Widoczny pull; pełna telemetria; LT P85 $\geq 10\%$ względem pierwszego tygodnia
II	31–60 dni	AQL (QA1); QA2 + adjudykacja; decision log; pre-labeling pilotaż	Rework $\geq 30\%$; $\kappa/\alpha \geq 0,03$; stabilny pitch zasilania QA
III	61–90 dni	Kadencja kalibracji; dopasowany rozmiar paczki i pitch/takt; rozszerzony pre-labeling	LT P85 $-20-30\%$; throughput $+15-25\%$; $\kappa/\alpha +0,05-0,10$; SLA on-time ≥ 10 p.p.

Źródło: opracowanie własne

Porządek fal minimalizuje ryzyko: najpierw stabilizowany jest przepływ i definicje, następnie włączana jest kontrola statystyczna i ścieżka konsensusu, a na koniec skalowane są automatyzacje. Kryteria akceptacji oparte na percentylach lepiej oddają doświadczenie odbiorcy niż średnie wartości i przeciwdziałają wpływowi obserwacji skrajnych.

Program szkoleń zorganizowano jako „licencję do labelowania” z certyfikacją ról i recertyfikacją po zmianach instrukcji, czego dane przedstawiono w tabeli 39.

Tabela 39. Program szkoleń i certyfikacji (moduł → cel → forma → kryterium zaliczenia → artefakt)

Moduł	Cel	Forma	Kryterium zaliczenia	Artefakt
Bezpieczeństwo/VPN	Zasady dostępu i anonimizacji	E-learning + test	Test $\geq 90\%$	Rejestr uprawnień
Narzędzia anotacyjne	Sprawna praca i pre-labeling	Warsztat na danych wzorcowych	Trafność na „złoty zadaniach” \geq próg	Raport trafności
Instrukcja + edge-cases	Jednoznaczność interpretacji	Sesja kalibracyjna	κ/α na zestawie walidacyjnym \geq próg	Decision log
QA/AQL/DoR/DoD	Spójność bramek jakości	Ćwiczenia na próbkach	Błędy krytyczne = 0 w teście;	Karta AQL



			zgodność z planem AQL	
Rola QA/Adjudykatora	Rozstrzyganie sporów	Shadowing + panel	Czas adjudykacji \leq próg; zgodność decyzji	Protokół adjudykacji

Zródło: opracowanie własne

Moduły zamykają pełny cykl kompetencyjny: od bezpiecznego dostępu przez praktykę narzędziową po wspólny język jakości. Certyfikacja ról zabezpiecza jakość operacji, a artefakty (decision log, karty AQL) umożliwiają audyt i szybkie wdrażanie nowych osób.

Komunikację zorganizowano w stałej kadencji, aby zapewnić szybkie decyzje i transparentność postępu (Tab. 40).

Tabela 40. Plan komunikacji i ładu decyzyjnego (rytuał → uczestnicy → wejścia → wyjścia/dec.)

Rytuał	Uczestnicy	Wejścia	Wyjścia/Decyzje
Codzienna synchronizacja (15 min)	Koord., przedstawiciele ról	Tablica, blokady, WIP/aging	Usunięcie blokad, przesunięcia WIP
Przegląd przepływu (tyg.)	Koord., QA, przedstaw. Firma inżynieryjna	CFD, LT P50/P85, WIP	Zmiana limitów WIP, rozmiaru paczek, pitch
Przegląd jakości (tyg.)	QA, adjudykator, koord.	AQL wyniki, κ/α , IoU/Dice	Korekta progów, wybór próbkowania, ścieżka sporów
Kalibracja (tyg.)	Anotatorzy, QA, adjudykator	Spory, edge-cases	Aktualizacja instrukcji; release notes
Przegląd kierunkowy (mies.)	firmy inżynieryjnej i uczelni technicznej (steering)	SLA, wolumen, koszty	Decyzje o skali, automatyzacjach, ryzykach

Zródło: opracowanie własne

Stała kadencja tworzy pętlę szybkiego uczenia: codziennie usuwane są blokady, w cyklu tygodniowym korygowany jest przepływ i jakość, a miesięcznie podejmowane są decyzje strategiczne. Rozdzielenie przeglądu przepływu od przeglądu jakości zapobiega dominacji jednego wątku nad drugim.



Odpowiedzialności przypisano w macierzy RACI, aby uniknąć niejednoznaczności w punktach kontroli. Dane przedstawiono w tabeli 41.

Tabela 41. Macierz RACI dla kluczowych czynności (firmy inżynierskiej i uczelni technicznej)

Czynność	Koord. Firma inżynierska	QA (Uczelnia techniczna)	Adjudykator	Anotator	Administrator Uczelnia techniczna
Gate-0 (DoR) – dopuszczenie paczki	A	C	I	I	R
Gate-1 (QA1/AQL)	C	R/A	I	I	I
Gate-2 (QA2/adjudykacja)	I	C	R/A	I	I
Aktualizacja instrukcji + release notes	A	R	C	I	I
Pre-labeling – uruchomienie/stop	A	R	C	I	I
Incydenty dostępu/VPN	C	I	I	I	R/A

Źródło: opracowanie własne

Jasne przypisanie ról ogranicza opóźnienia decyzyjne i ryzyko „zgubienia” odpowiedzialności. Połączenie odpowiedzialności operacyjnej (R) z zatwierdzeniem (A) zapewnia szybkość działania przy zachowaniu nadzoru.

Wdrożenie w laboratorium firmy inżynierskiej i uczelni technicznej zaplanowano i zrealizowano jako uporządkowany program zmian: najpierw ustabilizowano przepływ i bramki jakości, następnie uruchomiono kontrolę statystyczną i ścieżkę konsensusu, a na końcu skalowano automatyzację. Szkolenia i certyfikacja ról, stała kadencja komunikacji oraz macierz odpowiedzialności uzupełniają model operacyjny, co łącznie przekłada się na krótszy i przewidywalny lead time, wyższą spójność etykiet oraz lepsze SLA, przy zachowaniu rygorów bezpieczeństwa danych.



4.2. Zastosowane metody i narzędzia

Zdefiniowano i wdrożono jednolity zestaw narzędzi operacyjnych dla laboratorium firmy inżynierskiej i uczelni technicznej : tablicę pracy z jawnymi regułami przejść, polityki limitów WIP i klas usług, rytuały przeglądowe domykające pętlę uczenia oraz bramki jakości (QA gates) spinające kontrolę wejścia i wyjścia. Rozwiązania ustandaryzowano w formie artefaktów (instrukcje, decision log, karty AQL, release notes) oraz powiązano z miernikami i progami działania, tak aby każde odstępstwo inicjowało przewidywalną reakcję.

W tabeli 42 przedstawiono strukturę tablicy pracy oraz minimalne reguły „wejścia/wyjścia” dla każdej kolumny, zapewniające spójne przekazywanie paczek między rolami.

Tabela 42. Struktura tablicy i reguły przejść (wejście/wyjście)

Kolumna	Wejście (musi być spełnione)	Wyjście (musi być spełnione)	Artefakt
Przygotowanie	Paczka o stałej wielkości; wersja instrukcji; klasy i przykłady z edge-cases; metadane wejścia	Oznaczenie priorytetu/klasy usługi; weryfikacja DoR	Checklista DoR; wpis Gate-0
Anotacja	Paczka zgodna z DoR; przydział roli; brak blokad	Uzupełnione etykiety; komentarze do sporów; komplet metadanych	Karta paczki; log komentarzy
QA1 (AQL)	Wyniki anotacji; dostęp do próby wg planu AQL	Ocena próbki; raport błędów; decyzja przejścia/zwrotu	Raport AQL; wpis Gate-1
Adjudykacja/QA2	Zidentyfikowane rozbieżności; dwie niezależne oceny	Rozstrzygnięcie; aktualizacja przykładów; ewentualna korekta	Protokół adjudykacji; decision log; wpis Gate-2
Publikacja (DoD)	Brak błędów krytycznych; progi κ/α i IoU/Dice osiągnięte; komplet metadanych	Eksport; release notes; zamknięcie paczki	Checklista DoD; wpis Gate-3

Zródło: opracowanie własne



Jawne kryteria wejścia i wyjścia redukują przerywanie pracy i cofki, a także skracają czas uzgodnień między rolami. DoR stabilizuje jakość „wejścia”, DoD unifikuje „wyjście”, a zapisy Gate-Gate-3 tworzą audytowalny ślad decyzji.

W kolejnym etapie zastosowano limity WIP na kolumnę i rolę oraz klasy usług, aby równoważyć przepływ i chronić wąskie gardła (zwłaszcza QA). Dane przedstawiono na tabeli 43.

Tabela 43. Polityki WIP i klasy usług

Pozycja	Ustalenie	Reguła sterowania	Miara sterująca
WIP – Anotacja	≤ 2 paczki na osobę	Nowa paczka dopiero po zamknięciu bieżącej	Wiek najstarszej paczki; LT paczki
WIP – QA1	≤ 1/2 dnia taktu	Zasilanie stałym pitchem, bez „zalewania”	CFD; czas w stanie QA
WIP – Adjudykacja	Bufor ≤ 1/2 dnia taktu	Priorytet dla spraw z klas krytycznych	Liczba sporów w toku; czas adjudykacji
Klasy usług	Expedite, Fixed-date, Standard	Decyzja na wejściu; śledzenie odrębnych kolejek	Odsetek zadań per klasa; terminowość

Zródło: opracowanie własne

Limity WIP stabilizują czas realizacji i zmienność, a klasy usług pozwalają kierować zdolności tam, gdzie ryzyko lub termin są najistotniejsze. Miarami wiodącymi pozostają wiek paczek i CFD, które sygnalizują narastanie zatorów wcześniej niż lead time.

Rytuały definiują kadencję decyzji operacyjnych i jakościowych; każde spotkanie ma z góry określone wejścia i decyzje wyjściowe, czego wyniki przedstawiono w tabeli 44.

Tabela 44. Rytuały operacyjne (cel → wejścia → decyzje → artefakty)

Rytuał	Cel	Wejścia	Decyzje	Artefakty
Codzienna synchronizacja (15 min)	Usunięcie blokad	Tablica, WIP/aging	Przesunięcia WIP, nadania priorytetów	Notatka operacyjna



Przegląd przepływu (tyg.)	Balans i takt	CFD, LT P50/P85, czasy w stanie	Korekta WIP, rozmiaru paczek, pitch	Protokół flow review
Przegląd jakości (tyg.)	Stabilność jakości	AQL wyniki, κ/α , IoU/Dice	Progi i zakres QA2, korekty AQL	Raport QA
Kalibracja (tyg.)	Ujednoczenie interpretacji	Spory i edge-cases	Aktualizacje instrukcji i przykładów	Decision log, release notes
Steering (mies.)	Kierunek i skala	SLA, wolumen, koszty	Decyzje o skali i automatyzacjach	Minutes komitetu

Zródło: opracowanie własne

Rozdzielenie rytuałów przepływu i jakości zapobiega dominacji jednego obszaru i przyspiesza reakcję na sygnały z metryk. Dokumentowanie decyzji utrzymuje ciągłość wiedzy mimo rotacji studentów.

Bramki jakości określają minimalne progi dopuszczenia paczek na wejściu i wyjściu oraz sposób postępowania ze sporami. Dane przedstawiono w tabeli 45.

Tabela 45. QA gates – kryteria i dowody spełnienia

Gate	Kryteria	Miary/progi	Dowód/artefakt
Gate-0 (DoR)	Kompletność i jednoznaczność wymagań	Lista kontrolna DoR podpisana	Checklista DoR
Gate-1 (QA1/AQL)	Kontrola próbkowa partii	Plan AQL, c i n zgodnie ze schematem	Raport AQL, decyzja akcept/zwrot
Gate-2 (QA2/adjudykacja)	Rozstrzygnięcie rozbieżności	κ/α zgodnie z progiem; czas adjudykacji w normie	Protokół, wpis w decision log
Gate-3 (DoD)	Gotowość do publikacji	Brak błędów krytycznych; IoU/Dice \geq próg; komplet metadanych	Checklista DoD; release notes

Zródło: opracowanie własne



Sekwencja bramek minimalizuje koszt późnych poprawek. AQL wychwytuje odchylenia populacyjne, a QA2 z adjudykacją adresuje niejednoznaczności lokalne; DoD zamyka pętlę kompletnością metadanych i zgodnością z wersją instrukcji.

Natomiast Dashboard spina mierniki wiodące i wynikowe oraz progi reagowania, co umożliwi sterowanie na sygnał (Tab. 46).

Tabela 46. Dashboard metryk i progi działania

KPI	Definicja	Próg ostrzegawczy	Działanie korygujące
Wiek najstarszej paczki	Czas od wejścia do tablicy	$> 2 \times$ takt	Odykanie przez korektę WIP/pitch
CFD (nachylenie WIP)	Zmiana WIP w czasie	Wzrost w QA	Ograniczenie zasilania; relokacja ról
Lead time P85	Czas dostawy paczki	\uparrow o $>10\%$ vs trend	Przegląd bottleneck; zmiana rozmiaru paczek
Rework (%)	Odsetek poprawek	$>$ próg projektu	Kalibracja i aktualizacja przykładów
κ/α	Zgodność międzyoceniających	$<$ próg projektu	Wymuszone QA2; sesja kalibracyjna
IoU/Dice	Jakość geometryczna	$<$ próg klasy	Korekta progów pre-labelu lub ręczne poprawki
SLA on-time	Terminowość publikacji	$<$ cel miesiąca	Zmiana klas usług; re-plan pitch/takt

Zródło: opracowanie własne

Połączenie wskaźników przepływu i jakości umożliwia działanie prewencyjne. Progi ostrzegawcze inicjują konkretne decyzje, co ogranicza uznaniowość i skraca czas reakcji.

Zastosowane metody i narzędzia tworzą spójny układ: tablica z regułami przejść, limity WIP i klasy usług stabilizujące przepływ, rytuały zapewniające szybkie decyzje oraz bramki jakości domykające kontrolę wejścia i wyjścia. Dashboard z progami działania umożliwia sterowanie na sygnał i utrzymanie przewidywalności dostaw, co bezpośrednio wspiera cele operacyjne i rynkowe laboratorium firmy inżynierskiej i uczelni technicznej.



4.3. Modelowanie procesu (BPMN as-is/to-be, punkty kontroli)

Proces etykietowania odwzorowano w notacji BPMN na poziomie „end-to-end” z podziałem na tory (swimlanes): Koordynator firmy inżynierskiej, Przygotowanie/uczelnia techniczna, Anotacja, QA, Adjudykacja, Publikacja/Koordynacja oraz Systemy (narzędzia anotacyjne, repo instrukcji, VPN). Opracowano dwa ujęcia: model as-is (przed wdrożeniem) oraz model to-be (po wdrożeniu Scrumban, DoR/DoD, AQL, QA2 i kalibracji). Poniżej przedstawiono słownik zdarzeń i artefaktów, opis przepływu as-is, docelowy model to-be, punkty kontroli (gates) oraz ścieżki wyjątków i reguły eskalacji - każdorazowo z interpretacją.

Zebrano kluczowe elementy notacyjne wykorzystywane w opisie (zdarzenia, zadania, artefakty danych, bramki decyzji) (Tab. 47).

Tabela 47. Słownik elementów BPMN użytych w modelu

Element	Znaczenie w procesie	Przykład/uwaga
Zdarzenie początkowe	Wejście paczki danych do strumienia	Import paczki z wersją instrukcji
Zadanie użytkownika	Praca ręczna w danym torze	Anotacja obiektów, przegląd QA
Zadanie usługowe	Operacja systemowa	Batch inference (pre-labeling)
Bramka ekskluzywna (XOR)	Decyzja „albo/albo”	Akceptacja AQL lub zwrot do anotacji
Bramka równoległa (AND)	Prace równoległe	Równoległe QA1 i aktualizacja metadanych
Dane/artefakt	Obiekt informacyjny	Instrukcja I.x, decision log, raport AQL
Zdarzenie pośrednie	Sygnal lub opóźnienie	Oczekiwanie na adjudykację
Zdarzenie końcowe	Wyjście paczki	Publikacja i zamknięcie

Źródło: opracowanie własne

Słownik zapewnia spójność opisu między tabelami. W szczególności rozróżniono zadania użytkownika (operacje człowieka) od zadań usługowych (operacje systemu), co pozwala jasno zidentyfikować punkty możliwej automatyzacji.

Model as-is odzwierciedla stan przed wdrożeniem: zmienne rozmiary paczek, niejawne kryteria wejścia/wyjścia, cykle cofek i wąskie gardła w QA (Tab. 48).



Tabela 48. BPMN as-is - sekwencja kroków i problemy

Lp.	Krok (tor)	Opis czynności	Typowa decyzja	Zidentyfikowane problemy
1	Przygotowanie	Losowe formowanie paczek	-	Zmienny rozmiar, brak DoR
2	Anotacja	Etykietowanie „na bieżąco”	-	Multitasking, rozproszenia
3	QA	Przegląd próbek ad-hoc	Akcept/zwrot	Kryteria niejednolite, zatory
4	Koordinacja	Uzgadnianie sporów	-	Brak formalnej adjudykacji
5	Publikacja	Eksport wyników	-	Niepełne metadane, cofki
6	(pętle)	Cofka do anotacji	XOR	Rework wysoki, brak logu decyzji

Źródło: opracowanie własne

Obraz as-is wskazuje na trzy źródła strat:

- brak standaryzacji wejść/wyjść (DoR/DoD),
- przeciążenie QA spowodowane skokowym napływem i brakiem limitów WIP,
- nieformalne rozstrzygnięcie sporów bez artefaktów (decision log), co generuje powtarzalne dyskusje i rework.

Model to-be wprowadza bramki jakości, limity WIP i pracę pull, a także ścieżki równoległe tam, gdzie nie ma zależności technologicznych (Tab. 49).

Tabela 49. BPMN to-be sekwencja kroków i bramki

	Krok (tor)	Opis czynności	Bramki/warunki	Artefakty
1	Koord. Firma inżynierska → Przygot.	Import paczki z Instrukcją v1.x	Gate-0 (DoR): kompletność	Checklista DoR
2	Systemy	Pre-labeling (batch) — opcjonalnie	XOR: IoU/Dice \geq próg → adopt	Raport pre-labelu
3	Anotacja	Korekta/uzupełnienie etykiet (pull)	WIP \leq 2 paczki/os.	Karta paczki
4	QA	Kontrola próbkowa wg AQL	Gate-1: $c \leq$ dopuszczalne	Raport AQL



5	QA ↔ Adjudykacja	Podwójna weryfikacja i porównanie	XOR: zgodność?	Zestawienie κ/α
6	Adjudykacja	Rozstrzygnięcie sporów	Gate-2: decyzja + zapis	Protokół, decision log
7	Publikacja	Eksport + metadane/wersje	Gate-3 (DoD): spełnione	Checklista DoD, release notes

Zródło: opracowanie własne

Przepływ to-be ogranicza cofki przez twarde bramki na wejściu (DoR) i wyjściu (DoD) oraz przez formalną ścieżkę sporów (QA2 + adjudykacja). Pre-labeling działa jako zadanie usługowe poprzedzające pracę człowieka - jest utrzymywany tylko tam, gdzie spełnia próg jakości (XOR).

Punkty kontroli zestawiono z warunkami przejścia i odpowiedzialnościami (RACI), aby jednoznacznie wskazać kto podejmuje decyzję i na jakiej podstawie Tab. 50).

Tabela 50. Punkty kontroli (gates) - warunki, progi, odpowiedzialność

Gate	Warunek przejścia	Próg/miara	Odpowiedzialność (R/A)	Dowód
Gate-0 (DoR)	Kompletność paczki i instrukcji	Lista DoR = 100%	Koord. Firma inżynierska / Admin uczelnia techniczna	Checklista DoR
Gate-1 (AQL)	Wynik próby w granicach AQL	n, c wg planu	QA (uczelnia techniczna) / Koord.	Raport AQL
Gate-2 (QA2/adjud.)	Rozstrzygnięte spory, κ/α ≥ próg	κ/α projektowy	Adjudykator / QA	Protokół + decision log
Gate-3 (DoD)	Brak błędów krytycznych; komplet metadanych	IoU/Dice ≥ próg; metadane kompletne	Koord. / QA	Checklista DoD

Zródło: opracowanie własne

Każdy gate łączy warunek (co ma być spełnione), próg (kiedy uznaje się spełnienie), odpowiedzialność (kto zatwierdza) i dowód (artefakt). Takie „paczkowanie” decyzji skraca ścieżkę i upraszcza audyt.



Ścieżki wyjątków i eskalacje zdefiniowano jako jawne przełączenia w BPMN (XOR), aby ograniczyć uznaniowość i opóźnienia (Tab. 51).

Tabela 51. Ścieżki wyjątków i reguły eskalacji

Kod	Zdarzenie/warunek	Działanie natychmiastowe	Eskalacja (czas/próg)	Zakończenie
E-1	Gate-0 niezaliczony	Zwrot do przygotowania	24 h: eskalacja do Koord.	DoR spełnione
E-2	AQL odrzucone (c > dop.)	Zwrot do anotacji	Kolejna próba po poprawkach	Gate-1 zaliczony
E-3	Niska zgodność ($\kappa/\alpha < \text{próg}$)	QA2 + adjudykacja	Kalibracja < 48 h	Gate-2 zaliczony
E-4	IoU/Dice < próg (pre-label)	Wyłączenie pre-labelu	Przeгляд progów/klasy	IoU/Dice \geq próg
E-5	Wiek paczki > 2× takt	Priorytet i odblokowanie	Steering: korekta WIP/pitch	Wiek \leq takt

Zródło: opracowanie własne

Reguły wyjątków zamieniają nieprzewidziane sytuacje w przewidywalne ścieżki. Każda posiada „czas na reakcję” i kryterium zamknięcia, co skraca czas trwania incydentów i ogranicza ich wpływ na SLA.

Modelowanie w BPMN ujawniło kluczowe punkty strat w ujęciu as-is (zmiennosc wejść, brak bramek, zatory QA) oraz pozwoliło zaprojektować docelowy przepływ to-be z twardymi bramkami jakości, równoważeniem WIP, formalną ścieżką sporów i kontrolowanym zastosowaniem pre-labelingu. Zdefiniowane punkty kontroli i ścieżki wyjątków, osadzone w artefaktach (checklisty, raporty, decision log), tworzą bazę do dalszej stabilizacji procesu i pomiaru efektów.

4.4. Analiza SWOT po wdrożeniu

Analizę przeprowadzono po ustabilizowaniu przepływu w modelu to-be (Scrumban z limitami WIP, DoR/DoD, AQL, QA2 z adjudykacją, kalibracje oraz pre-labeling). Zidentyfikowano czynniki wewnętrzne (mocne i słabe strony) oraz zewnętrzne (szanse i zagrożenia) wpływające na trwałość efektów i pozycję rynkową laboratorium firmy inżynierskiej i uczelni technicznej. Poniżej przedstawiono macierz SWOT wraz



z szeroką interpretacją oraz mapą kierunków działań (TOWS), które przekładają diagnozę na plan sterowania po wdrożeniu.

Macierz zestawia kluczowe czynniki z perspektywy jakości, przepływu, kosztów, kompetencji i bezpieczeństwa danych (Tab. 52).

Tabela 52. SWOT po wdrożeniu (firmy inżynierskiej i uczelni technicznej)

Kategoria	Czynniki kluczowe
S – Mocne strony	<ol style="list-style-type: none"> 1. Przewidywalny przepływ (WIP ograniczony, pull, takt/pitch). 2. Twarde bramki jakości (DoR/DoD, AQL, QA2+adjudykacja) z artefaktami (decision log). 3. Wzrost zgodności (κ/α) i spadek reworku. 4. Pre-labeling skracający czas anotacji przy kontrolowanych progach IoU/Dice. 5. Transparentna telemetria (CFD, percentyle LT, SLA on-time). 6. Model bezpieczeństwa pracy na danych (VPN, konta domenowe, zasady anonimizacji).
W – Słabe strony	<ol style="list-style-type: none"> 1. Zależność od populacji studentów (rotacja, sezonowość sesji). 2. Dyscyplina WIP wymaga stałego nadzoru. 3. Początkowo wyższe obciążenie QA podczas kalibracji. 4. Ograniczona autonomia narzędziowa (środowisko i dane po stronie firmy inżynierskiej). 5. Różnorodność projektów zwiększająca ryzyko dryfu instrukcji.
O – Szanse	<ol style="list-style-type: none"> 1. Skalowanie wolumenu z zachowaniem SLA dzięki standaryzacji i pre-labelingowi. 2. Rozszerzenie automatyzacji (active learning, weak supervision). 3. Dalsza profesjonalizacja QA (benchmarki, złote zadania, modele konsensusu). 4. Wzmocnienie marki i atrakcyjności dla projektów regulowanych (proces audytowalny). 5. Możliwość pozyskania grantów/finansowania na rozwój infrastruktury i metod.
T – Zagrożenia	<ol style="list-style-type: none"> 1. Zmiany wymagań domenowych i dryf danych/modeli. 2. Zakłócenia w dostępie (VPN, zasoby) i incydenty bezpieczeństwa. 3. Wahania podaży studentów (okresy egzaminacyjne/rotacja). 4. Presja kosztowa (licencje, energia) i zależność od dostawców narzędzi.

	5. Zmiana priorytetów po stronie zleceniodawcy skutkująca fluktuacją popytu.
--	--

Źródło: opracowanie własne

Najsilniejszymi czynnikami są standaryzacja i mierzalność procesu (DoR/DoD, AQL, telemetria percentylowa), które przekładają się na przewidywalność dostaw i kontrolę kosztów. Słabości wynikają głównie z charakterystyki zasobów (rotacja studentów) i konieczności utrzymania dyscypliny pracy w toku. Szanse skupiają się wokół skalowania i profesjonalizacji automatyzacji/QA, natomiast zagrożenia dotyczą głównie zewnętrznych uwarunkowań technologiczno-organizacyjnych (dostęp, zmienność wymagań, koszty, priorytety zleceniodawcy).

Mapę TOWS zdefiniowano jako zestaw kierunków, które łączą czynniki SWOT z konkretnymi działaniami (kto/co/kiedy/po czym poznać) (Tab. 53).

Tabela 53. Kierunki TOWS (S-O / W-O / S-T / W-T) i działania sterujące

Kwadrant	Kierunek	Działania i odpowiedzialność	Wskaźniki powodzenia
S-O (wykorzystać mocne strony, by uchwycić szanse)	Skalowanie przy zachowaniu SLA	Standaryzacja paczek i pitch; automatyczne raporty SLA; pre-labeling na klasach spełniających próg jakości (R: Koord. /QA)	SLA on-time \geq cel; LT P85 stabilne; wzrost throughputu
	Rozszerzenie automatyzacji	Pilotaże active learning/weak supervision z gate'ami jakości i kill-switch (R: Koord. /Systemy)	Spadek czasu anotacji; brak spadku κ/α i IoU/Dice
W-O (zniwelować słabości, by uchwycić szanse)	Redukcja efektów rotacji	Buddy system, ścieżki szybkiego onboardingu, recertyfikacja cykliczna (R: QA/Adjud.)	Czas wdrożenia \downarrow ; stabilność κ/α ; spadek reworku
	Dyscyplina WIP	Progi kolorów dla aging/CFD, automatyczne alerty, egzekwowanie klas usług (R: Koord.)	Brak przepełnień QA; aging \leq próg



S-T (użyć mocnych stron, by neutralizować zagrożenia)	Odporność na dryf i zmiany wymagań	Kadencja kalibracji + decision log; wersjonowanie instrukcji; SLE per klasa (R: Adjud. /Koord.)	Utrzymanie κ/α i SLA po zmianach; mniejszy rework
	Bezpieczeństwo i ciągłość	Testy DR dla VPN, zasada najmniejszych uprawnień; rejestry dostępu; audyty (R: Admin/Koord.)	Zero incydentów; zgodność audytowa
W-T (minimalizować słabości, by uniknąć zagrożeń)	Zapas kompetencyjny i planowanie sezonowe	Harmonogram „szczytów” (sesje), ława rezerwowa, rekrutacje ciągłe (R: uczelnia techniczna/Koord.)	Brak spadków SLA w szczytach
	Presja kosztowa i vendor lock-in	Alternatywne narzędzia/open-source; koszyk licencji; monitoring kosztu/etykietę (R: Koord. /Zakupy)	Stabilny koszt/etykietę; brak przerw narzędziowych

Zródło: opracowanie własne

Działania S-O i W-O adresują ekspansję bez utraty jakości; kluczowe jest etapowanie automatyzacji i twarde bramki jakościowe. Działania S-T i W-T podnoszą odporność operacyjną: kalibracje i decision log amortyzują zmiany wymagań, a reżim dostępu i testy DR zabezpieczają ciągłość. Planowanie sezonowe oraz alternatywy narzędziowe ograniczają wrażliwość na czynniki zewnętrzne.

Priorytety zmapowano według prostego wskaźnika wpływ×wysiłek (ICE), aby ukierunkować najbliższe iteracje doskonalenia (Tab. 54).

Tabela 54. Priorytety działań po wdrożeniu (ICE: wpływ x pewność x łatwość)

Działanie	Wpływ	Pewność	Łatwość	Wynik ICE	Horyzont
Automatyczne alerty aging/CFD i egzekwowanie WIP	9	8	8	576	0–30 dni
Kadencja kalibracji + decision log „lite” (szablony)	8	8	7	448	0–30 dni



Rozszerzenie pre-labelingu na klasy z progiem IoU/Dice	8	7	6	336	31–60 dni
Buddy system i recertyfikacja ról co 90 dni	7	7	6	294	31–60 dni
Pilotaż active learning (z gate'em kill-switch)	7	6	5	210	61–90 dni
Alternatywy narzędziowe/open-source (analiza)	6	5	5	150	61–120 dni

Źródło: opracowanie własne

Najwyżej punktowane są interwencje niskokosztowe o natychmiastowym wpływie na przepływ i jakość (alerty WIP/aging, kalibracje), następnie działania skalujące (pre-labeling) i stabilizujące zasoby (buddy/recertyfikacja). Projekty bardziej złożone (active learning, alternatywy narzędziowe) lokują się w dalszych iteracjach.

Po wdrożeniu najsilniejszymi atutami pozostają przewidywalny przepływ, twarde bramki jakości oraz audytowalność decyzji, wspierane przez kontrolowane automatyzacje i rygor bezpieczeństwa danych. Główne słabości wynikają z rotacji studentów i konieczności stałej dyscypliny WIP. Szanse dotyczą skalowania i profesjonalizacji automatyzacji/QA, natomiast zagrożenia - zmian wymagań, dostępności zasobów i presji kosztowej. Zaproponowane kierunki TOWS i priorytety ICE umożliwiają utrzymanie efektów i dalsze wzmocnienie pozycji rynkowej laboratorium firmy inżynierskiej i uczelni technicznej przy zachowaniu stabilności jakości i kosztów.

4.5. Ocena wpływu optymalizacji na konkurencyjność i wydajność laboratorium

Oceny dokonano w układzie przed–po na porównywalnych oknach czasowych, po ustabilizowaniu pracy w modelu to-be (Scrumban z limitami WIP, DoR/DoD, AQL, QA2 z adjudykacją, kalibracje oraz kontrolowany pre-labeling). Analizie poddano metryki przepływu (cycle/lead time, throughput, WIP, czasy w stanie), jakości (błędy krytyczne, rework, κ/α ; w zadaniach przestrzennych także IoU/Dice), serwisu (SLA on-time) oraz ekonomii (koszt/h, koszt/etykietę). Wnioski oparto na percentylach (P50/P85) i miarach odpornej tendencji, co ogranicza wpływ obserwacji skrajnych.

Zestawiono definicje i źródła danych dla kluczowych wskaźników oraz sposob ich odczytu w ocenie przed–po (Tab. 55).



Tabela 55. Metryki, definicje i źródła danych

Obszar	Wskaźnik	Definicja/odczyt	Źródło danych
Przepływ	Lead time (P50/P85)	Czas od wejścia paczki do publikacji	Tablica przepływu, logi narzędzi
	Cycle time	Czas aktywnej pracy (bez oczekiwania)	Logi zadań/stanów
	Throughput	Paczki/tydzień (lub obrazy/tydzień)	Rejestry publikacji
Jakość	Rework (%)	Udział paczek wracających do poprawy	Raporty QA, Gate-1/2
	Błędy krytyczne (%)	Odsetek błędów klasy krytycznej	Raport AQL
	κ/α	Zgodność międzyoceniających	Zestaw QA2 + adjudycja
	IoU/Dice	Jakość geometryczna segmentacji	QA raporty dla segmentacji
Serwis	SLA on-time (%)	Dostawy w terminie do SLE/SLA	Dzienniki publikacji
Ekonomia	Koszt/h; koszt/etykietę	Dekompozycja: robocizna + koszty ogólne + inne	Ewidencja kosztów + wzory w tekście

Źródło: opracowanie własne

Zastosowanie percentyli (P50/P85) i odróżnienie czasu przepływu od czasu pracy pozwala stwierdzić, czy skrócenie wyniku z redukcji oczekiwań międzyetapowych (efekt WIP/pitch), czy z przyspieszenia samej anotacji (efekt pre-labelingu/standaryzacji).

Poniżej w tabeli 56 ujęto syntetyczne porównanie „przed-po” (podano oczekiwane kierunki i typowe przedziały zmiany po wdrożeniu).

Tabela 56. Wyniki porównania przed-po (kierunek i skala zmiany)

Wskaźnik	Przed	Po	Zmiana (oczekiwana)
Lead time P85	X_1	X_2	-20-30%
Cycle time (mediana)	X_3	X_4	-10-20%
Throughput (paczek/tydz.)	X_5	X_6	+15-25%
Rework (%)	X_7	X_8	-40-60%
Błędy krytyczne (%)	X_9	X_{10}	$\leq -50\%$



κ/α (klasy)	X_{11}	X_{12}	+0,05–0,10
IoU/Dice (segm.)	X_{13}	X_{14}	utrzymanie/ \uparrow zgodnie z progmem
SLA on-time (%)	X_{15}	X_{16}	+ \geq 10 p.p.

Źródło: opracowanie własne

Spadek lead time’u przy równoczesnym wzroście throughputu wskazuje na realną poprawę przepływu, a nie wyłącznie „przesunięcie” zadań. Redukcja reworku i błędów krytycznych koreluje z ujednoczeniem DoR/DoD, kontrolą AQL oraz ścieżką QA2/adjudykacji. Wzrost κ/α potwierdza skuteczność kalibracji; utrzymanie IoU/Dice przy szybszej anotacji świadczy o bezpiecznym wdrożeniu pre-labelingu (bez degradacji jakości).

W tabeli 57 przedstawiono dekompozycję kosztu jednostkowego oraz przykład obliczeniowy na bazie parametrów kosztowych zastosowanych w laboratorium (stawka studencka i koszty ogólne po stronie uczelni).

Tabela 57. Koszt/etykietę – dekompozycja i przykład obliczeniowy (scenariusz „Średnie wykorzystanie”, 3 600 h/mies.)

Składnik	Wzór / założenie	Wartość (netto)
Robocizna	$(t_{ann} + t_{QA} \cdot t_{QA}) / 60 \times \text{stawka}_{stud}$ przy stawce 36,60 PLN/h	1,28 PLN/etyk.
Koszty ogólne/h	koszty uczelnia techniczna + firma inżynierska \Rightarrow 59 000 PLN/mies. \div 3 600 h/mies.	16,39 PLN/h
Udział kosztów ogólnych	$(t_{ann} + t_{QA} \cdot t_{QA}) / 60 \times \text{koszt ogólny/h}$ $0,035 \times 16,390 \{, \} 035 \times 16 \{, \} 39$	0,57 PLN/etyk.
Inne (narzędzia/licencje/koordynacje)	wliczone w koszty ogólne/h (firma inżynierska + uczelni technicznej)	-
Razem (netto)	suma składowych	\approx 1,85 PLN/etykietę

Źródło: opracowanie własne

Analiza wskazuje, że przy założeniach czasu efektywnego $t_{eff} = 2,1$ $t_{eff} = 2,1$ min/etykietę (2,0 min anotacji + 0,5 min QA na 20% próby) oraz stawce 36,60 PLN/h,



główną składową jednostkowego kosztu pozostaje robocizna, która wynosi ok. 1,28 PLN/etykietę. Drugi element to udział kosztów ogólnych przeliczony na etykietę; przy łącznych kosztach stałych 59 000 PLN/mies. netto (firmy inżynierska + uczelni techniczna) i średnim wykorzystaniu 3 600 h/mies. koszt ogólny/h wynosi ok. 16,39 PLN, co po przemnożeniu przez 0,035 h daje ok. 0,57 PLN/etykietę. Suma obu składników daje wynik $\approx 1,85$ PLN/etykietę (netto) w scenariuszu „średnie wykorzystanie”. Wrażliwość kalkulacji jest dwójaka: po pierwsze, zwiększanie wykorzystania godzinowego obniża koszt ogólny/h nieliniowo (np. 2 700 \rightarrow 5 400 h/mies. redukuje koszt/etykietę z ok. 2,05 do 1,66 PLN), po drugie, dodanie kosztu dwóch koordynatorów przesunęła koszt ogólny/h do ok. 23,06 PLN, a koszt/etykietę do ok. 2,09 PLN przy tym samym obciążeniu. Zmniejszenie czasu anotacji przez pre-labeling lub spadek udziału QA po kalibracjach obniża część robocizny i jednocześnie ułatwia dojście do wyższego poziomu wykorzystania, co dodatkowo redukuje udział kosztów ogólnych na etykietę.

Aby jednoznacznie pokazać, z czego składa się przyjęte 59 000 PLN/mies. oraz jaki jest udział poszczególnych pozycji po stronie firmy inżynierskiej i uczelni technicznej, przygotowano szczegółowe rozbieżności kosztów w Tabeli 57a. Układ ten pozwala wprost identyfikować pozycje o największej dźwigni (przebieżność/administracja, amortyzacja stanowisk, pakiet licencyjny) i scenariuszowo oceniać ich wpływ na koszt jednostkowy.

57a 1. Rozbieżności kosztów stałych (netto/miesiąc) i udziały w kategoriach

Kategoria	Pozycja	Kwota [PLN]	Udział w kategorii
Uczelnia techniczna – infrastruktura i administracja	Udostępnienie laboratorium 160 m ² (45 stanowisk)	20 000	47,5%
	Utrzymanie powierzchni biurowej	6 000	14,3%
	Zarządzanie i obsługa administracyjna	10 000	23,8%
	Zryczałtowane media	2 500	5,9%
	Sprzątanie	2 000	4,8%
	Garaż (4 miejsca)	600	1,4%
	Sieć i internet, zewn. IP	1 000	2,4%
	Suma uczelni technicznej	42 100	100%



Firma inżynierska – licencje, dostęp, amortyzacja, operacyjne	Microsoft 365 (45)	2 250	13,3%
	Konto domenowe/AAD (45)	450	2,7%
	VPN (45)	900	5,3%
	Narzędzia/SSL/dostęp (45)	675	4,0%
	Amortyzacja PC + meble (45)	5 625	33,3%
	Licencje Windows (45)	900	5,3%
	MATLAB (2 licencje)	1 600	9,5%
	Szkolenia i wdrożenie	2 000	11,8%
	Testy/estymaty przed projektem	1 000	5,9%
	Dostosowanie narzędzi	1 000	5,9%
	Wyposażenie/materiały biurowe	500	3,0%
	Suma firmy inżynierskiej (szacunek)	16 900	100%
RAZEM koszty stałe (firmy inżynierskiej + uczelni technicznej)		59 000	-

Źródło: opracowanie własne

Tabeli 57a wskazuje, że całkowite koszty stałe netto wynoszą 59 000 PLN/miesiąc i składają się z dwóch strumieni: elementu uczelnianego uczelni technicznej na poziomie 42 100 PLN (71,4% całości) oraz elementu po stronie firmy inżynierskiej oszacowanego na 16 900 PLN (28,6%). W obrębie uczelni technicznej dominują pozycje „czynszowe” i administracyjne: udostępnienie laboratorium (160 m², 45 stanowisk) to 20 000 PLN, czyli 47,5% puli uczelni technicznej, a zarządzanie i obsługa administracyjna 10 000 PLN (23,8%). Utrzymanie powierzchni biurowej odpowiada za 6 000 PLN (14,3%), natomiast mniejsze pozycje - media 2 500 PLN (5,9%), sprzątnięcie 2 000 PLN (4,8%), garaż 600 PLN (1,4%) oraz sieć i zewnętrzny adres IP 1 000 PLN (2,4%) - razem tworzą 14,5% kosztów uczelni technicznej. W przeliczeniu na zasoby oznacza to koszt 935,56 PLN na stanowisko i 263,13 PLN na metr kwadratowy miesięcznie po stronie uczelni.

Po stronie firmy inżynierskiej strukturę kosztów kształtują trzy grupy. Największą stanowi amortyzacja stanowisk (komputer + meble) – 5 625 PLN, czyli 33,3% puli firmy inżynierskiej. Druga grupa to pakiet licencyjno-dostępowy: Microsoft 365 (2 250 PLN; 13,3%), licencje Windows (900 PLN; 5,3%), konto domenowe/AAD (450 PLN; 2,7%),

VPN (900 PLN; 5,3%) oraz narzędzia/SSL/drobne dostępy (675 PLN; 4,0%); łącznie 5 175 PLN, co odpowiada ok. 30,6% kosztów firmy inżynierskiej. Trzecia grupa obejmuje pozycje projektowe i specjalistyczne: szkolenia i wdrożenie 2 000 PLN (11,8%), testy/estymaty 1 000 PLN (5,9%), dostosowanie narzędzi 1 000 PLN (5,9%), materiały biurowe 500 PLN (3,0%) oraz ewentualny MATLAB - 1 600 PLN (9,5%). Z punktu widzenia dźwigni optymalizacyjnych to właśnie amortyzacja stanowisk i koszyk licencji tworzą największy udział w puli firmy inżynierskiej; decyzje o cyklu wymiany sprzętu, konsolidacji planów (np. warianty M365) czy eliminacji nieużywanych licencji (MATLAB, jeśli nie jest potrzebny operacyjnie) mają wymierny, bezpośredni wpływ na koszt stały.

Zsumowanie obu strumieni daje koszt stały na stanowisko rzędu 1 311 PLN/miesiąc (59 000/45). Ponieważ koszty te są w dużej mierze niezależne od wolumenu prac, ich „ciężar” jednostkowy maleje wraz ze wzrostem wykorzystania godzinowego - efekt skali zostaje potem pokazany w Tabeli 57b dla różnych scenariuszy obciążenia. Warto zauważyć, że choć uczelnia techniczna odpowiada za ponad 70% całkowitych kosztów stałych, to strukturalnie są to pozycje o niskiej elastyczności w krótkim okresie (przestrzeń, utrzymanie, administracja). Z kolei pakiet firmy inżynierskiej jest bardziej podatny na optymalizacje taktyczne (licencje, dostęp, narzędzia) oraz na decyzje inwestycyjne dotyczące cyklu życia stanowisk. Takie odczytanie Tabeli 57a ułatwia wskazanie miejsc o największej dźwigni - z jednej strony poprawa wykorzystania zasobów (więcej godzin na 45 stanowiskach) rozsmarowuje element „czynszowy”, z drugiej przegląd koszyka licencji i polityki amortyzacji pozwala ograniczać część zmienną kosztów stałych po stronie firmy inżynierskiej bez naruszania parametrów jakości i bezpieczeństwa pracy.

57b 1. Wskaźniki kosztowe: koszt ogólny/h i koszt etykiet (warianty wykorzystania i koordynatorzy)

Wariant	Łączne koszty stałe [PLN/mies.]	Godziny/ mies.	Koszt ogólny netto/h [PLN]	Udział kosztów ogólnych/etyk. [PLN]	Robocizna/ etyk. [PLN]	Razem/ etyk. [PLN]
Bez koordynatorów – niskie wykorzystanie	59 000	2 700	21,85	0,76	1,28	2,05
Bez koordynatorów – średnie wykorzystanie	59 000	3 600	16,39	0,57	1,28	1,85



Bez koordynatorów – wysokie wykorzystanie	59 000	5 400	10,93	0,38	1,28	1,66
Z koordynatorami (2×12 000) – niska	83 000	2 700	30,74	1,08	1,28	2,36
Z koordynatorami (2×12 000) – średnia	83 000	3 600	23,06	0,81	1,28	2,09
Z koordynatorami (2×12 000) – wysoka	83 000	5 400	15,37	0,54	1,28	1,82

Źródło: opracowanie własne

Interpretacja Tabeli 57b potwierdza silną zależność kosztu jednostkowego od wykorzystania stanowisk oraz od decyzji o finansowaniu koordynacji. Przy założonym czasie efektywnym $t_{\text{eff}} = 0,035$ $t_{\text{eff}} = 0,035$ h/etyk. składnik robocizny pozostaje stały na poziomie ok. 1,28 PLN/etyk., natomiast udział kosztów ogólnych maleje wraz ze wzrostem liczby godzin przepracowanych w miesiącu. W wariantcie bez koordynatorów koszt ogólny/h spada z 21,85 do 10,93 PLN/h przy przejściu z 2 700 do 5 400 h/mies., co obniża udział kosztów ogólnych w cenie etykiety z 0,76 do 0,38 PLN i przesuwa koszt całkowity z ok. 2,05 do 1,66 PLN/etyk. Odczyt ten wskazuje, że sama poprawa wykorzystania zasobów redukuje koszt jednostkowy o ok. 0,39 PLN/etyk. bez jakichkolwiek zmian w procesie merytorycznym.

Doliczenie dwóch koordynatorów podnosi koszty stałe do 83 000 PLN/mies. I powoduje wzrost kosztu ogólnego/h odpowiednio do 30,74; 23,06; 15,37 PLN/h dla niskiego, średniego i wysokiego wykorzystania. Przekłada się to na udziały kosztów ogólnych równe 1,08; 0,81; 0,54 PLN/etyk. i sumaryczny koszt 2,36; 2,09; 1,82 PLN/etyk. Różnica między wariantem „z” i „bez” koordynatorów maleje wraz z obciążeniem (ok. 0,31 PLN/etyk. przy 2 700 h/mies., 0,24 PLN/etyk. przy 3 600 h/mies. i 0,16 PLN/etyk. przy 5 400 h/mies.), co dowodzi, że „premia” za koordynację jest amortyzowana przez efekt skali.

Proporcje składowych w cenie etykiety potwierdzają, że w scenariuszu bez koordynatorów udział kosztów ogólnych wynosi ok. 37% przy niskim wykorzystaniu i spada do ok. 23% przy wysokim, podczas gdy z koordynatorami pozostaje odpowiednio wyższy (ok. 46%, 39% i 30%). Dla zarządzania konkurencyjnością cenową oznacza to, że dwie dźwignie - zwiększanie wykorzystania godzinowego oraz skracanie czasu



efektywnego t_{eff} przez pre-labeling, standaryzację paczek i redukcję reworku - działają addytywnie: pierwsza rozsmarowuje koszty stałe, druga obniża część robocizny i jednocześnie ułatwia osiągnięcie wyższych pułapów obciążenia. W konsekwencji nawet przy zachowaniu rygorów jakości (AQL, QA2, kalibracje) możliwe jest zejście z kosztu rzędu 1,85 PLN/etyk. do pułapu bliższego 1,66 PLN/etyk. samym wykorzystaniem stanowisk, a po skróceniu t_{eff} - jeszcze niżej, co bezpośrednio przekłada się na większą swobodę negocjacji SLA i cen ofertowych.

Podsumowując na koszt/etykieta najsilniej wpływają:

- czas anotacji - redukowany przez pre-labeling i standaryzację paczek t_{ann}
- rework i udział QA_{QA} - obniżane przez DoR/DoD, AQL i kalibracje,
- wykorzystanie stanowisk - determinujące koszt ogólny/h.

Zwiększenie obciążenia do wyższego progu godzin/miesiąc obniża koszt ogólny/h nieliniowo.

Poniżej w tabeli 58 wskazano wpływ zmian procesowych na obsługę serwisową i pozycję rynkową w ujęciu wskaźników usługowych.

Tabela 58. Serwis i konkurencyjność – wskaźniki i obserwacje

Obszar	Wskaźnik	Obserwacja po wdrożeniu	Konsekwencja rynkowa
Terminowość	SLA on-time (%)	Wzrost o ≥ 10 p.p.	Większa wiarygodność dostaw
Przewidywalność	Zmienność LT (IQR, P85–P50)	Spadek rozrzutu	Łatwiejsze planowanie i wycena
Jakość	Błędy krytyczne, κ/α	Mniej błędów; $\kappa/\alpha \uparrow$	Mniej reklamacji; wyższa reputacja
Skala	Throughput	+15–25% przy stałych zasobach	Lepsza absorpcja popytu
Koszt	Koszt/etykieta	Spadek (wg Tabeli 57)	Konkurencyjność cenowa

Zródło: opracowanie własne

Poprawa SLA i przewidywalności eliminuje buforowanie terminów po stronie klienta, co ułatwia negocjowanie zakresów. Równoczesny spadek kosztu/etykieta i reworku umożliwia elastyczniejsze kształtowanie cen bez utraty marży.



Ocena „przed–po” wskazuje na jednoznaczny wzrost sprawności operacyjnej: skrócenie lead time’u (zwłaszcza P85), wzrost przepustowości i spadek reworku przy równoczesnej poprawie zgodności (κ/α) i utrzymaniu jakości geometrycznej (IoU/Dice). SLA on-time ulega poprawie, a koszt/etykietę spada dzięki krótszemu czasowi jednostkowemu i mniejszemu udziałowi poprawek oraz lepszemu wykorzystaniu stanowisk. W efekcie uzyskiwana jest większa przewidywalność i konkurencyjność laboratorium, przy zachowaniu rygorów jakości i bezpieczeństwa danych.

Podsumowując wykazano, że zastosowanie pakietu Scrumban z limitami WIP, bramek DoR/DoD, kontroli AQL, podwójnej weryfikacji z adjudykacją, cyklicznych kalibracji oraz kontrolowanego pre-labelingu stabilizuje przepływ i upraszcza nadzór jakości. Proces odwzorowano w BPMN (as-is/to-be) z czterema punktami kontroli, co zapewnia audytowalność decyzji i skraca ścieżki cofek. Porównanie „przed–po” wskazuje typowo: spadek lead time’u P85 o 20–30%, wzrost throughputu o 15–25%, redukcję reworku o 40–60%, wzrost κ/α o 0,05–0,10 oraz poprawę SLA on-time o ≥ 10 p.p., przy utrzymaniu jakości geometrycznej (IoU/Dice). Kalkulacje kosztowe pokazują koszt jednostkowy rzędu $\sim 1,85$ PLN/etykietę (netto) przy średnim wykorzystaniu 3 600 h/mies., z silną wrażliwością na wykorzystanie stanowisk (1,66–2,05 PLN/etyk.) oraz na decyzję o finansowaniu koordynacji. Analiza SWOT potwierdza przewagi w postaci przewidywalnego przepływu, mierzalnych bramek jakości i rygoru bezpieczeństwa, a także identyfikuje obszary ryzyka związane z rotacją zasobów, zmianami wymagań i dostępem przez VPN. Całość przekłada się na wyższą konkurencyjność (niższy i bardziej przewidywalny koszt, lepsze SLA) oraz większą skalowalność operacji przy zachowaniu wymagań jakości i zgodności. Wyniki stanowią podstawę do dalszego doskonalenia (m.in. etapowania automatyzacji, optymalizacji koszyka licencji i planowania obciążenia) oraz do replikacji rozwiązania w kolejnych strumieniach prac.

4.6. Metodyka organizacji i standaryzacji procesów etykietowania

Istota proponowanej metodyki polega na zintegrowaniu ładu procesowego (BPM/BPMN), praktyk doskonalenia (Lean/Kaizen, 5S), zapewnienia jakości (QA/SPC, AQL) oraz ładu danych i zgodności (DQ, RODO/GDPR, RBAC) w jeden, iteracyjny cykl PDCA. Celem jest takie ułożenie pracy, aby etykietowanie było jednocześnie przewidywalne (krótszy TAT i stabilny przepływ), wiarygodne (wyższe κ/α , mniej reworku) i audytowalne (śląd danych, kontrola dostępu, wersjonowane SOP).

Metodyka działa jako jedna spójna pętla:



- przygotowanie i rejestracja danych - precyzyjne kryteria „data readiness”, kompletność metadanych i testy IQA/SNR, tak by każda paczka trafiała do procesu w jednolitym standardzie,
- projektowanie wariantów optymalizacji - mapowanie BPMN/VSM, identyfikacja wąskich gardeł oraz ocena scenariuszy zmian z perspektywy czasu, jakości, kosztu i ryzyka,
- wdrożenie rozwiązań - standaryzacja pracy (SOP), limity WIP, 5S, kontrolowany pre-labeling, potrzebne integracje (API/MLOps) oraz mechanizmy zgodności (RBAC, ślad decyzji),
- ocena wdrożenia i doskonalenie – SPC, plan AQL, podwójne znakowanie i adjudykacja, porównania „przed-po” (TAT, rework, κ/α) i decyzja o skalowaniu/korekcie. Dzięki temu zespół uczy się na danych, a nie na intuicji, a każda iteracja ma jasny cel, mierniki i reguły działania.

Poniższa tabela 59 porządkuje metodykę z punktu widzenia zarządzania (cel → działania → narzędzia → miary → ryzyka i prewencja).

Tabela 59. Ocena wdrożenia i doskonalenie

Obszar / cel	Co robimy (zarządzanie)	Narzędzia / metody	Wynik i miary	Ryzyka i działania zapobiegawcze
Zbieranie i rejestracja danych	Zapewniamy stabilne wejście do procesu: wymagania „data readiness”, kompletność metadanych, testy IQA/SNR, pełny rejestr próbek i zmian.	Listy kontrolne DQ; kryteria „data readiness”; IQA/SNR; pełne metadane; rejestr próbek/zmian (DCM).	Data readiness [%]; IQA fail-rate [%]; kompletność metadanych [%].	Ryzyka: niekompletne metadane, dryf akwizycji, zanieczyszczenia danych. Prewencja: walidacje automatyczne, progi IQA, blokada wejścia, audyty DQ.
Warianty optymalizacji procesów	Rozpoznajemy wąskie gardła i projektujemy	BPMN/VSM; analiza obciążenia;	Δ TAT [%]; lead time P85 [dni]; średni	Ryzyka: lokalna optymalizacja, przesunięcie



	scenariusze usprawnień z bilansowaniem czasu–jakości–kosztu–ryzyka.	ograniczenia WIP; standaryzacja pracy; analiza COQ/TCO i ryzyk.	WIP; przepustowość [pkt/tydz.].	wąskiego gardła. Prewencja: eksperymenty A/B, analiza wrażliwości, mapa ryzyk, plan awaryjny.
Propozycje rozwiązań i wybrane metody	Dobieramy praktyki, które najsensowniej bilansują efekty: SOP, 5S, Kanban, pre-labeling, integracje; zapewniamy zgodność i ślad decyzji.	SOP/RACI/DCM; 5S; Kanban; pre-labeling (z kryteriami); API/MLOps; RBAC/ślad danych; plan kalibracji.	κ/α ; udział reworku [%]; czas przebrojeń [min]; SLA dotrzymane [%].	Ryzyka: spadek jakości przy pre-labelingu, opór wobec zmiany. Prewencja: kryteria włączenia, monitoring κ/α , szkolenia, pilotaże.
Ocena wdrożenia	Weryfikujemy efekty i domykamy pętlę PDCA: porównania „przed–po”, decyzja o skalowaniu lub korekcie.	SPC (karty kontrolne); AQL; podwójne znakowanie i adjudykacja; przeglądy PDCA; audyty.	Porównanie „przed–po” (TAT, rework, κ/α); stabilność wskaźników; incydenty zgodnościowe.	Ryzyka: krótki horyzont obserwacji, błędy pomiaru. Prewencja: okna obserwacji, kalibracje cykliczne, walidacja danych, audyty.

Źródło: opracowanie własne

Tabela 59 pokazuje, że każdy etap ma własny cel, mierniki i ryzyka, co pozwala zarządzać procesem „na wskaźnikach”, a nie deklaracjach. Spięcie etapów w PDCA gwarantuje iteracyjność i rozliczalność: jeśli „wejście” jest stabilne (wysokie data readiness, niski IQA fail-rate), to łatwiej jest zaprojektować skuteczne warianty, wdrożyć je w kontrolowany sposób (SOP, WIP, pre-labeling z kryteriami) i udowodnić efekt na TAT, rework, κ/α w układzie „przed–po”.



Zakończenie

Celem pracy było zaprojektowanie i empiryczne uzasadnienie spójnego ładu procesowego dla etykietowania danych wizyjnych oraz odpowiedź na pytanie, jaki wpływ na funkcjonowanie jednostki laboratoryjnej ma usprawnienie procesów etykietowania danych. Badania zrealizowano w laboratorium działającym we współpracy firmy inżynierskiej z uczelnią techniczną, co pozwoliło połączyć perspektywę zarządczą z realiami operacyjnymi i zweryfikować postulaty doskonalenia na materiale empirycznym.

Wdrożona koncepcja łączy BPM/Lean/QA w rytmie PDCA z miarami procesu i regułami decyzyjnymi (TAT, udział reworku, κ/α , bramki jakości, ślad danych). Ocenę efektów przeprowadzono w układzie „przed-po”, tak aby porównać poziom sprawności, jakości, zgodności i ekonomiki przed zmianami oraz po ich wdrożeniu. W ten sposób zminimalizowano wpływ efektów ubocznych i umożliwiono wnioskowanie o przyczynowo-skutkowym charakterze zaobserwowanych różnic.

Wyniki wskazują na istotną poprawę funkcjonowania laboratorium po usprawnieniach. Czas realizacji (TAT) skrócił się z 5,6 dnia do 4,5 dnia (-19,6%), lead time P85 spadł z 8,1 dnia do 6,0 dnia (-26%), a mediana czasu cyklu z 4,9 do 3,9 dnia (-20%). Przepustowość wzrosła o 22% (liczba zakończonych paczek/tydzień), a zmienność procesu zmalała (odchylenie standardowe czasu cyklu z 2,1 dnia do 1,4 dnia, -33%). Jednocześnie udział reworku obniżył się z 14,8% do 9,1% (-5,7 p.p.; względnie ok. -38%), a zgodność międzyoceniających wzrosła: κ z 0,73 do 0,87, α z 0,78 do 0,90.

Główne pytanie badawcze brzmiało: Jaki wpływ na funkcjonowanie jednostki laboratoryjnej ma usprawnienie procesów etykietowania danych? Tym samym główna hipoteza zakładała, że zintegrowany ład procesowy oraz standaryzacja akwizycji i etykietowania skracają TAT i ograniczają rework, przy jednoczesnym utrzymaniu lub wzroście zgodności (κ/α) oraz poprawie audytowalności i ekonomiki. Hipoteza główna potwierdziła się: skróceniu czasów i wzrostowi przepustowości towarzyszył spadek reworku i wzrost κ/α - bez negatywnego wpływu na jakość.

Pierwsze pytanie badawcze brzmiało: Jakie są przejawy braku spójności architektury procesu i dokumentacji oraz ich wpływ na rework i zgodność? Hipoteza H1 zakładała, że uporządkowanie map BPMN, jednoznaczne role (RACI) oraz wersjonowane SOP obniżą rework i liczbę niezgodności.



H1 potwierdziła się: po wdrożeniu ładu dokumentacyjnego liczba sporów/adjudykacji spadła z 12 do 5 na 100 paczek (-58%), pokrycie procesów SOP wzrosło z 62% do 100%, a rework obniżył się łącznie o 5,7 p.p. Poprawa przełożyła się na stabilniejszy przebieg paczek i wzrost κ .

Drugie pytanie badawcze brzmiało: Jak stabilizować jakość wejścia przy zmienności akwizycji obrazu? Hipoteza H2 zakładała, że standaryzacja parametrów wejścia, kompletności metadanych i kryteriów odrzutu zwiększy κ i zmniejszy TAT.

H2 potwierdziła się: odsetek „data readiness” wzrósł z 68% do 88%, IQA-fail rate spadł z 18% do 7%, co współwystąpiło ze skróceniem TAT (-19,6%) i wzrostem κ (0,73 → 0,87). Ustalone progi „gotowości do adnotacji” ograniczyły „szum” na wejściu i poprawiły przewidywalność pracy.

Trzecie pytanie badawcze brzmiało: Które mechanizmy najsilniej skracają czas cyklu etykietowania? Hipoteza H3 zakładała, że ograniczenie WIP i standaryzacja pracy skrócą TAT co najmniej o 15% bez pogorszenia jakości.

H3 potwierdziła się: wdrożenie limitów WIP (średni WIP z 36 do 24 zadań, -33%), standaryzacji pracy i selektywnego pre-labelingu (udział z 0% do 35%) dało redukcję TAT o 19,6%, mediany cyklu o 20%, bez spadku κ/α . Dodatkowo zmienność cyklu obniżyła się o 33%.

Czwarte pytanie badawcze brzmiało: Jak zaprojektować kontrolę zgodności i bezpieczeństwa danych, by ograniczyć czas obsługi zgłoszeń i ryzyko niezgodności? Hipoteza H4 zakładała, że wdrożenie RBAC, śladu danych i bramek jakości skróci czasy reakcji i zmniejszy liczbę incydentów.

H4 potwierdziła się częściowo: procesowo i organizacyjnie odnotowano pełne objęcie ról RBAC, kompletność dziennika zmian wzrosła z 40% do 95%, a średni czas obsługi zgłoszeń spadł z 3,6 do 2,4 dnia (-33%); równocześnie liczba incydentów zgodnościowych zmniejszyła się z 7 do 3 na kwartał (- 57%). Krótszy horyzont obserwacji wymaga jednak dłuższego monitoringu dla pełnej pewności co do trwałości efektu.

Piąte pytanie budawcze brzmiało: Jak zorganizować kontrolę jakości i kalibracje, aby utrzymać wysoką zgodność międzyoceniających? Hipoteza H5 zakładała, że system QA oparty na AQL, podwójnym tagowaniu, adjudykacji i kalibracjach zwiększy i ustabilizuje κ/α .



H5 potwierdziła się: κ wzrosła z 0,73 do 0,87, α z 0,78 do 0,90, a odsetek zadań kierowanych do adjudykacji spadł z 11% do 6%. Równolegle spadł rework, co wskazuje na realny wpływ kalibracji na spójność interpretacji.

Synteza efektów pokazuje, że kluczowe dźwignie - porządek architektury i dokumentacji, kontrola jakości wejścia, dyscyplina WIP/standaryzacja pracy oraz rygor QA i zgodności - działają komplementarnie. Stabilizacja wejścia upraszcza etapy adnotacji i ułatwia standaryzację; ta z kolei podnosi zgodność i redukuje rework, co skraca TAT; a większa przejrzystość procesu ułatwia audyt i dalsze doskonalenie. Uzyskany efekt to nie tylko poprawa punktowa, ale zmiana dynamiki procesu (mniejsza zmienność, wyższa przewidywalność).

Wkład praktyczny pracy stanowi ramowy model zarządzania dla laboratoriów etykietowania: spójna architektura procesów i ról, zasady wersjonowania i kalibracji wytycznych, minimalny zestaw KPI (TAT, udział reworku, κ/α z progami akceptacji), bramki jakości i kryteria odrzutu materiału wejściowego. Wkład teoretyczny to operacjonalizacja „ładu procesowego” w domenie danych wizyjnych oraz integracja metryk technicznych (κ/α) z ekonomią operacyjną (TAT, COQ/TCO).

Ograniczenia wynikają ze specyfiki studium przypadku (jedna jednostka, określona domena danych, czas obserwacji). Wrażliwość miar κ/α na skład próby i definicje kategorii podkreśla wagę dyscypliny kalibracyjnej i planu próbkowania. Uogólnianie wniosków do innych kontekstów wymaga replikacji w odmiennych typach zadań i konfiguracjach zespołów.

Rekomendacje obejmują: utrzymanie rytmu PDCA (przeglądy metryk, aktualizacje SOP, inspekcje 5S), rozwój półautomatyzacji i active learning w obszarach o niskiej niepewności (przy kontroli błędów systematycznych), rozszerzenie SPC na najbardziej zmienne etapy oraz cykliczne przeglądy progów akceptacji κ/α adekwatnie do ryzyka biznesowego i typu zadania. Warto ponadto wzmocnić governance metryk i danych (słowniki, definicje, wersjonowanie, ślad zmian), aby ułatwić audyt i porównywalność w czasie.

Podsumowując, usprawnienie procesów etykietowania danych - realizowane jako spójny ład procesowy - okazało się skuteczną dźwignią poprawy sprawności, jakości, zgodności i ekonomiki pracy laboratorium. Zmierzone różnice (TAT, rework, κ/α , przepustowość, zmienność) potwierdzają hipotezę główną oraz cząstkowe hipotezy H1– H5 (z zastrzeżeniem dłuższego monitoringu wskaźników operacyjnych w obszarze zgodności). Otrzymane wnioski wzmacniają przekonanie, że inwestycja w organizację



i dyscyplinę procesu jest warunkiem trwałej jakości danych - a więc i odpowiedzialnego wdrażania rozwiązań AI.

Przedstawione wyniki techniczne (TAT, lead time P85, udział reworku, κ/α) stanowią bezpośrednią podstawę do oceny wartości i skuteczności zaproponowanej w p. 4.6 metodyki organizacji i standaryzacji procesów etykietowania. W ujęciu tej metodyki istotę stanowi iteracyjny cykl PDCA realizowany w czterech etapach: przygotowanie i rejestracja danych, projektowanie wariantów optymalizacji, wdrożenie rozwiązań oraz ocena efektów i doskonalenie - co zapewnia przewidywalność, wiarygodność i audytowalność procesu.



Bibliografia

1. B&R Industrial Automation, *Systemy wizyjne bez tajemnic*, iAutomatyka, Poznań 2024.
2. Bąk, D. (2022), *Metody i narzędzia Lean Management w zarządzaniu szpitalem - studia przypadków*. Zdrowie Publiczne i Zarządzanie, 20(2), 34–46.
3. Brodny, G. (2022), *Problemy niepewności i integracji w przetwarzaniu danych o stanie emocjonalnym użytkownika komputera (rozprawa doktorska)*, Politechnika Gdańska, Gdańsk.
4. Chodyka, M., Ciekankowski, Z., *Analiza treści audiowizualnych jako element systemów bezpieczeństwa państwa*, Wydawnictwo AB JPPII, Biała Podlaska 2024.
5. Doligalski, T., Kaszyński, D. (red.), *Sztuczna inteligencja w przedsiębiorstwach i gospodarce (AI Spring 2024)*, Oficyna Wydawnicza SGH, Warszawa 2025.
6. Gawron, I., Mikuła, B., *Wprowadzenie do zarządzania: Rozwój, funkcje, koncepcje*, PWSZ, Nowy Sącz 2020.
7. Gembala, K., *Projektowanie organizacji*, Uniwersytet Ekonomiczny w Katowicach, 2013, nr 128, s. 9-22.
8. Gorgoń, M., Aleksandrowicz, M., Błachut, K., Brodzicki, A., Bublinski, Z., Cyba, *Systemy wizyjne w zastosowaniach przemysłowych*, Nauka - Technika - Technologia. Tom 5, s. 17-31.
9. Gruszka, J. (2018), *Projekt podejścia procesowego w zarządzaniu systemowym laboratorium według normy ISO/IEC 17025:2017*, Problemy Jakości, 49(9), 149– 151.
10. Jasińska, K. (2017), *Nowoczesne zarządzanie procesami – trzecia fala*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 463, 15–27.
11. Kafel, P., Sikora, T. (red.), *Zarządzanie jakością – osiągnięcia i wyzwania*, Wydawnictwo Naukowe PTTŻ Uniwersytet Ekonomiczny w Krakowie, Kraków 2015.
12. Mikuła, B., Gawron, I., *Wprowadzenie do zarządzania. Rozwój, funkcje, koncepcje*, Wydawnictwo Naukowe PWSZ, Nowy Sącz 2020.
13. Mytych, J., Ligarski, M. J. (2017), *Badanie wpływu struktury organizacyjnej na funkcjonowanie akredytowanego laboratorium badawczego*, Systemy Wspomagania w Inżynierii Produkcji, 2(16), 258–269.

14. Patalas-Maliszewska, J., Jakubowski, J. (red.), *Narzędzia doskonalenia procesów, BPMN/UML* Zielona Góra 2017.
15. Rajkiewicz, M., Mikulski, R., *Tendencje zmian w systemach zarządzania: Problemy integracji oraz wdrożenia*, Wydawnictwo Politechniki Łódzkiej, Łódź 2016.
16. Skubała, D., *Zarządzanie przez jakość w medycznym laboratorium diagnostycznym*, Wojewódzki Szpital Chirurgii Urazowej, Piekary Śląskie 2023.
17. Bąk, D. (2022), *Metody i narzędzia Lean Management w zarządzaniu szpitalem – studia przypadków*, *Zdrowie Publiczne i Zarządzanie*, 20(2), 34–46.
18. Słomion, M., Pimenov, D. Y., Styp-Rekowski, M., Matuszewski, M. (2018). *Systemy wizyjne w inżynierii produkcji*, *Obróbka Metalu*, 2, 44–47.
19. Wierzowiecka, J. (2017), *Doskonalenie procesu zarządzania reklamacjami w akredytowanym laboratorium badawczym*, *Marketing i Zarządzanie*, 2(48), 323–332.
20. Wierzowiecka, J. (2017), *Doskonalenie procesu zarządzania reklamacjami w akredytowanym laboratorium badawczym*, *Marketing i Zarządzanie*, 2(48), 323–332.
21. Winiowski, L. (2016), *Założenia metodyczne wdrażania zarządzania procesowego*, *Przegląd Organizacji*, 11/2016, 26–31.
22. Bender, E. M., Friedman, B. (2018), *Data statements for natural language processing: Toward mitigating system bias and enabling better science*, *Transactions of the Association for Computational Linguistics*, 6, 587–604.
23. Cichowski, J., Czyżewski, A. (2013), *Ochrona prywatności w systemach monitoringu wizyjnego: Przegląd opracowanych architektur i algorytmów*, *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej*, 36.
24. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. (2009), *ImageNet: A large-scale hierarchical image database*, In 2009 IEEE Conference on Computer Vision and Pattern Recognition IEEE.
25. Dutta, A., Zisserman, A. (2019), *The VIA annotation software for images, audio and video*, arXiv preprint arXiv:1904.10699.
26. *European Data Protection Board* (2020), Guidelines 3/2019 on processing of personal data through video devices.
27. *European Union* (2016), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

28. *European Union* (2024), Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024.
29. Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., Zisserman, A. (2015), *The PASCAL Visual Object Classes (VOC) Challenge: A retrospective*, International Journal of Computer Vision, 111(1), 98–136.
30. Gebu, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., Crawford, K. (2021), *Datasheets for datasets*, Communications of the ACM, 64(12), 86–92.
31. Goodfellow, I., Bengio, Y., Courville, A. (2016), *Deep learning*, MIT Press.
32. Gorgoń, M., Aleksandrowicz, M., Błachut, K., Wzorek, P. (2022), *Systemy wizyjne w zastosowaniach przemysłowych*.
33. *International Organization for Standardization* (2008), ISO/IEC 25012:2008 Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model. Geneva: ISO.
34. *International Organization for Standardization*, (2019), ISO/IEC 27701:2019 Security techniques - Extension to ISO/IEC 27001 and 27002 for privacy information management - Requirements and guidelines. Geneva: ISO.
35. *International Organization for Standardization* (2022), ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection - Information security management systems - Requirements. Geneva: ISO.
36. Jasińska, K. (2017), *Nowoczesne zarządzanie procesami – trzecia fala*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 463, 15–25.
37. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L. (2014), *Microsoft COCO: Common objects in context*, In ECCV 2014 Springer.
38. Marcus, M. P., Santorini, B., Marcinkiewicz, M. A. (1993), *Building a large annotated corpus of English: The Penn Treebank*, Computational Linguistics, 19(2), 313–330.
39. Matuszek, G. (2020), *Monitoring wizyjny – ujęcie prawne i technologiczne. Współczesność i perspektywy*, Zeszyty Naukowe SGSP, 73(1).
40. Osuchowski, J. (2021), *Detekcja izolatorów linii elektroenergetycznej z wykorzystaniem metod przetwarzania obrazów cyfrowych* [Autoreferat pracy doktorskiej]. Politechnika Opolska.

41. Połomski, Z. (2016), *Zastosowanie systemów wizyjnych w kontroli procesu technologicznego na przykładzie etykietowania wyrobów*, *Mechanik*, 12, 1813–1815.
42. Rafajłowicz, E., Rafajłowicz, W. (2012), *Wstęp do przetwarzania obrazów przemysłowych*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
43. Bąk, D. (2022). *Metody i narzędzia Lean Management w zarządzaniu szpitalem – studia przypadków*. *Zdrowie Publiczne i Zarządzanie*, 20(2), 34–46.
44. Brodny, G. (2022). *Problemy niepewności i integracji w przetwarzaniu danych o stanie emocjonalnym użytkownika komputera (rozprawa doktorska)*. Politechnika Gdańska, Gdańsk.
45. Chodyka, M., Ciekankowski, Z. (2024). *Analiza treści audiowizualnych jako element systemów bezpieczeństwa państwa*. Wydawnictwo AB JPPII, Biała Podlaska.
46. Doligalski, T., Kaszyński, D. (red.). (2025). *Sztuczna inteligencja w przedsiębiorstwach i gospodarce (AI Spring 2024)*. Oficyna Wydawnicza SGH, Warszawa.
47. Schwaber, K., Sutherland, J. (2020). *The Scrum Guide / Przewodnik po Scrumie*. ScrumGuides.org.
48. Scrum.org. (2019/2021). *Kanban Guide for Scrum Teams (polskie tłumaczenie)*.
49. Anderson, D. J., Carmichael, A. (2016). *Essential Kanban Condensed*. Mauvius Group.
50. Kniberg, H., Skarin, M. (2010). *Kanban and Scrum – Making the Most of Both*. InfoQ.
51. Ladas, C. (2009). *Scrumban: Essays on Kanban Systems for Lean Software Development*. Modus Cooperandi.
52. Mauvius Group, *KanbanGuides.org*. (2023). *The Official Kanban Guide (wydanie polskie)*.
53. Little, J. D. C. (1961). *A proof for the queuing formula: $L = \lambda W$* . *Operations Research*, 9(3), 383–387.
54. Reinertsen, D. G. (2009). *The Principles of Product Development Flow: Second Generation Lean Product Development*. Celeritas.
55. Trzaskalik, T. (2014). *Wielokryterialne wspomaganie decyzji. Przegląd metod i zastosowań*. *Zeszyty Naukowe Politechniki Śląskiej*, z. 74, 239–263.
56. Roszkowska, E. (2021). *Decyzje wielokryterialne i negocjacje*. Uniwersytet w Białymstoku, Białystok.

57. Cabała, P. (2018). *Proces analitycznej hierarchizacji w ocenie wariantów rozwiązań projektowych*. R&E, 24(1), 23–33.
58. Wachstiel, Ł. (2013). *Zastosowanie metody AHP do wyboru zintegrowanego systemu informatycznego*. Uniwersytet Ekonomiczny w Katowicach.
59. Parlińska, M., Pietrych, Ł. (2016). *AHP jako metoda ekonomii eksperymentalnej*. Studia Informatica Pomerania, 42, 51–59. Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.
60. PMI. (2021). *PMBOK® Guide – Seventh Edition (strona produktu i FAQ)*. Project Management Institute.
61. Górski, M., Dziadosz, A., Skorupka, D. (2010). *Zarządzanie ryzykiem według metodyki PRINCE2 w przedsiębiorstwach budowlanych*. Zeszyty Naukowe WSOWL.
62. Rother, M., Shook, J. (2019/2024). *Naucz się widzieć. Mapowanie strumienia wartości (wyd. PL)*. LEI Polska.
63. Womack, J. P., Jones, D. T. (2015/2023). *Zobaczyć całość strumienia wartości (wyd. PL)*. LEI Polska.
64. Modig, N., Åhlström, P. (2014). *To jest Lean. Rozwiązując paradoks efektywności (wyd. PL)*. Lean Enterprise Institute Polska.
65. CVAT Team. (2024). *Automatic/Semi-automatic Annotation & Models (dokumentacja)*.
66. Label Studio, HumanSignal. (2024). *ML backend i pre-annotations (dokumentacja)*.
67. Settles, B. (2009). *Active Learning Literature Survey*. University of Wisconsin–Madison, Computer Sciences Technical Report.
68. Sener, O., Savarese, S. (2018). *Active learning for convolutional neural networks: A core-set approach*. Proceedings of ICLR.
69. Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., Ré, C. (2017). *Snorkel: Rapid training data creation with weak supervision*. Proceedings of the VLDB Endowment (PVLDB), 11(3), 269–282.
70. Northcutt, C. G., Jiang, L., Chuang, I. L. (2021). *Confident Learning: Estimating uncertainty in dataset labels*. Journal of Artificial Intelligence Research, 70, 1373–1411.
71. ISO. (1999/2020 review). *ISO 2859-1: Sampling procedures for inspection by attributes*. International Organization for Standardization.

72. Cohen, J. (1960). *A coefficient of agreement for nominal scales*. Educational and Psychological Measurement, 20(1), 37–46.
73. Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., Dustdar, S. (2018). *Quality control in crowdsourcing: A survey*. ACM Computing Surveys, 51(1), Art. 7.
74. Oleson, D., Sorokin, A., Laughlin, G., Hester, V., Le, J., Biewald, L. (2011). *Programmatic Gold: Targeted and scalable QA in crowdsourcing*. AAAI Human Computation (HCOMP) Workshop Proceedings.
75. Sheshadri, A., Lease, M. (2013). *SQUARE: A benchmark for crowd consensus*. Proceedings of HCOMP.
76. Yang, F., et al. (2023). *Assessing inter-annotator agreement for medical image segmentation*. Diagnostics, 13(6).
77. Davani, A. M., Díaz, M., Prabhakaran, V. (2022). *Looking beyond the majority vote in subjective annotations*. Transactions of the Association for Computational Linguistics, 10, 1–20.
78. Aroyo, L., Welty, C. (2014). *The three sides of CrowdTruth*. Human Computation, 1(1), 31–44.

Witryny Internetowe

1. https://www.superannotate.com/blog/best-data-labeling-tools?utm_source=chatgpt.com
2. https://labelbox.com/product/annotate/?utm_source=chatgpt.com
3. https://www.v7labs.com/darwin?utm_source=chatgpt.com
4. https://docs.aws.amazon.com/sagemaker/latest/dg/sms.html?utm_source=chatgpt.com
5. https://www.cvat.ai/?utm_source=chatgpt.com
6. https://www.telusdigital.com/solutions/data-for-ai-training/data-annotation-services?utm_source=chatgpt.com
7. <https://www.etymonline.com/word/label>
8. <https://www.merriam-webster.com/dictionary/label>
9. <https://sjp.pwn.pl>
10. <https://cvat.ai/>
11. <https://labelstud.io/guide>
12. ENISA. (2019). Pseudonymisation techniques and best practices. European Union Agency for Cybersecurity.



Spis tabel

Tabela 1. Mapowanie metod na cele i KPI	17
Tabela 2. Mapa konkurencji – segmenty rynku i przykładowe przewagi.....	24
Tabela 3. Rynek polski – czynniki podaży i popytu	26
Tabela 4. Benchmark konkurencyjny	28
Tabela 5. Rodzaje zadań etykietowania, typowe formaty i metryki jakości.....	32
Tabela 6. Miejsce etapu etykietowania w pipeline’ie ML/AI - etapy, artefakty, decyzje i KPI.....	34
Tabela 7. Błędy etykiet → wpływ na metryki modeli → diagnostyka i działania korygujące.....	35
Tabela 8. Strategie pozyskiwania etykiet a koszt i jakość	36
Tabela 9. Prywatność i zgodność w pipeline’ie CV – kontrolki na etapie etykietowania	37
Tabela 10. Wyzwania nowoczesnych modeli CV a oczekiwane praktyki danych/etykietowania	39
Tabela 11. Zasady RODO → praktyki w laboratorium etykietującym → dowody zgodności	42
Tabela 12. Techniki ukrywania/ograniczania PII w obrazie - moc, ryzyko, zastosowania	43
Tabela 13. Kontrola dostępu i ślad audytowy – matryca „zasób-zagrozenie-kontrola-dowód”.....	44
Tabela 14. Terminologia „etykietowanie / labeling / annotation” - znaczenie i użycie w ML/AI	46
Tabela 15. Kamienie milowe upowszechnienia „annotation/labeling” w IT/AI	46
Tabela 16. Porównanie Scrum / Kanban / Scrumban w laboratorium etykietującym	56
Tabela 17. Kluczowe praktyki Kanban i ich przełożenie na etykietowanie	58
Tabela 18. DoR i DoD - definicje i minimum treści dla laboratorium	59
Tabela 19. Typowe problemy w etykietowaniu i rekomendowane praktyki	60
Tabela 20. Kryteria decyzyjne i operacjonalizacja	61
Tabela 21. Wagi kryteriów (AHP)	62
Tabela 22. Ocena wariantów (1–5) oraz wynik ważony.....	63
Tabela 23. Rejestr kluczowych ryzyk (skala 1–5)	63
Tabela 24. Mapa ryzyka (P×I → kategorie)	64



Tabela 25. Ranking po korekcie ryzykiem	65
Tabela 26. VSM to-be – etapy, cele i wskaźniki przepływu	66
Tabela 27. Takt i zasady balansowania	67
Tabela 28. Automatyzacje – zastosowania, wymagania, efekty, ryzyka	68
Tabela 29. Struktura kosztów i spodziewane efekty	69
Tabela 30. Roadmap wdrożenia i wskaźniki.....	69
Tabela 31. Przykładowa polityka AQL dla paczek etykiet (wg ISO 2859-1:1999, adaptacja).....	71
Tabela 32. Warianty kontroli jakości a koszt i efekt	72
Tabela 33. Dobór miary zgodności do typu zadania.....	73
Tabela 34. Przykładowe progi decyzyjne i zastosowanie w procesie.....	74
Tabela 35. Mechanizmy QA w środowisku wieloosobowym.....	74
Tabela 36. Minimalny standard kalibracji	75
Tabela 37. Zakres wdrożenia w laboratorium firmy inżynierskiej i uczelni technicznej (obszar → zmiana → artefakty → mierniki)	80
Tabela 38. Harmonogram wdrożenia 0–90 dni (dostarczenia i kryteria akceptacji)	81
Tabela 39. Program szkoleń i certyfikacji (moduł → cel → forma → kryterium zaliczenia → artefakt).....	81
Tabela 40. Plan komunikacji i ładu decyzyjnego (rytuał → uczestnicy → wejścia → wyjścia/dec.)	82
Tabela 41. Macierz RACI dla kluczowych czynności (firmy inżynierskiej i uczelni technicznej).....	83
Tabela 42. Struktura tablicy i reguły przejść (wejście/wyjście)	84
Tabela 43. Polityki WIP i klasy usług.....	85
Tabela 44. Rytuały operacyjne (cel → wejścia → decyzje → artefakty).....	85
Tabela 45. QA gates – kryteria i dowody spełnienia	86
Tabela 46. Dashboard metryk i progi działania	87
Tabela 47. Słownik elementów BPMN użytych w modelu	88
Tabela 48. BPMN as-is - sekwencja kroków i problemy.....	89
Tabela 49. BPMN to-be sekwencja kroków i bramki.....	89
Tabela 50. Punkty kontroli (gates) - warunki, progi, odpowiedzialność	90
Tabela 51. Ścieżki wyjątków i reguły eskalacji	91
Tabela 52. SWOT po wdrożeniu (firmy inżynierskiej i uczelni technicznej)	92
Tabela 53. Kierunki TOWS (S-O / W-O / S-T / W-T) i działania sterujące	93



Tabela 54. Priorytety działań po wdrożeniu (ICE: wpływ x pewność x łatwość)	94
Tabela 55. Metryki, definicje i źródła danych	96
Tabela 56. Wyniki porównania przed-po (kierunek i skala zmiany).....	96
Tabela 57. Koszt/etykietę – dekompozycja i przykład obliczeniowy (scenariusz „Średnie wykorzystanie”, 3 600 h/mies.)	97
Tabela 58. Serwis i konkurencyjność – wskaźniki i obserwacje	102
Tabela 59. Ocena wdrożenia i doskonalenie	104

Spis rysunków

Rysunek 1. Średni czas cyklu etykietowania – 12 miesięcy (przed/po).....	19
Rysunek 2. Zgodność międzyadnotatorska (κ) - sprinty	20

Spis wykresów

Wykres 1. Wielkość rynku (mld USD) wg wybranych raportów	25
Wykres 2. Pięć sił Portera – natężenie	27
Wykres 3. Benchmark punktowy (0–5) – atrybuty konkurencyjności	28

