Tag 5 · Das erste Iteration Review · Gruppe 2

Ziel
Eure Entscheidung stützt sich auf geprüfte Befunde aus zwei Recherchewellen und einen getesteten Ablaufprototyp.

Fertig, wenn
die Entscheidung Belege und Schwelle nennt, die zweite Welle einen geprüften Befund oder eine begründete Grenze liefert und eine Änderung der Arbeitsweise feststeht.

Gruppenübung – Nachmittag

1 Befunde prüfen

Vergleicht eure geprüfte Belegtabelle mit der vorab gesetzten Schwelle. Trennt belastbare Messwerte, Gegenbelege und offene Übertragbarkeit.

Ergebnis: 29 Datenpunkte, Quellenart je Quelle: 5 unabhängig/wissenschaftlich, 11 Anbieterangaben mit eigener Metrik, 6 Praxisberichte, 3 Amts-/Standard-, 4 Rechts-/Behördenquellen. Belastbar: 14 % vollständig im Self-Service (Gartner, n = 5.728), 16,3 % DACH-Median (140 Shops / 2,6 Mio. Tickets), OBI 20 %. Gegenbelege: Alibaba-Feldexperiment (680.676 Chats, −0,858 Punkte; aufgestaute Eskalation +40,8 % Zeit), Containment-Artefakte (Abbrecher, Close-on-silence, Reopen). Offen: Übertragbarkeit auf unseren Shop.

2 Entscheiden

Euer Product Owner entscheidet über Weiterführen, Anpassen oder Stoppen des untersuchten Ansatzes und nennt die tragenden Belege. Prüft den Ablaufprototyp gegen Sprint Goal und Definition of Done.

Ergebnis: Entscheidung ANPASSEN (Begründung rechts). Tragende Belege: 14 % und 16,3 % widerlegen 60 % als Gesamtquote; 61,2 % Top-Quartil und die Sprünge 1,8 %→~60 % bzw. 25 %→~75 % nach Systemanbindung stützen 60 % nur für einen engen Intent-Block. Ablaufprototyp: Test 1 und Test 2 bestanden; Definition of Done erfüllt — offen bleibt die Messung des Trigger-Recalls.

3 Zweite Welle schneiden

Wählt die entscheidende Beleglücke oder einen Widerspruch aus und formuliert gezielte Folgeaufträge mit benötigtem Beleg. Nutzt dafür denselben Rechercheprompt mit den neuen Aufträgen.

Ergebnis: Lücken ausgewählt: (a) „Übergabe und Qualitätsprüfung“ war nicht eigenständig abgedeckt, (b) kein empirischer Schwellenwert für Übergabe-Trigger, (c) Intent-Verteilung des eigenen Bestands unbekannt. Folgeauftrag A: Übergabe-Design und laufende QA — Beleg: Designprinzipien + False-Containment-Metriken. Folgeauftrag B: Intent-Tagging im eigenen Bestand + Trigger-Precision/-Recall — Beleg: eigene Messung.

4 Nachrecherchieren

Führt die Folgeaufträge über denselben geprüften Ausführungsweg aus und öffnet die neuen Originalquellen. Bleibt ein Beleg unauffindbar, begründet die Grenze und entwerft den nötigen Pilotversuch.

Ergebnis: Folgeauftrag A ausgeführt, Originalquellen geöffnet (HICSS 2021, DSK-OH 2024, Platforma 2026): Die Übergabe ist die schwächste Stelle — nur 10 % erleben sie als reibungslos, 59 % müssen ihr Anliegen erneut erklären; keiner dieser Fehler senkt die Containment-Quote, mehrere erhöhen sie. Folgeauftrag B: kein öffentlicher Beleg auffindbar → Grenze begründet, als Pilotversuch entworfen (n ≥ 400, stratifiziert, Doppelcodierung mit κ).

5 Anpassen

Ergänzt die Belegtabelle und bewertet die Schwelle erneut. Euer Product Owner aktualisiert die Entscheidung und ordnet die nächsten Items; das Team passt die erste Scheibe an, soweit der Befund sie betrifft.

Ergebnis: Belegtabelle um 6 Belege erweitert (Übergabe/QA, Recht, Trigger). Schwelle neu bewertet: 60 % bleiben als Best-Case-Wert für einen engen Intent-Block mit Live-Datenanbindung haltbar, als Gesamtquote nicht. Erste Scheibe angepasst: proaktive Übergabe nach mehr als drei erfolglosen Vorschlägen oder Zeitgrenze; die Quote wird nur zusammen mit Trigger-Recall und Recontact Rate (72 h) gewertet.

6 Arbeitsweise verbessern

Haltet eine konkrete Änderung aus der Retrospective auf der Gruppenfläche fest. Legt bei weitergeführtem Ansatz den nächsten Prüf- oder Umsetzungsschritt fest.

Ergebnis: Änderung: Der Zugriffsweg wird vor dem Start je Subagent fest vorgegeben (Reader/curl + pdftotext statt Suchtreffer), weil das Extraktionswerkzeug funktionslos war. Nächster Schritt: Pilotmessung im eigenen Shop — n ≥ 400 Fälle, Metrik vor dem Lauf einfrieren, untere Wilson-Grenze ≥ 60 % als Nachweiskriterium.

GRUNDLAGE AUS DEM VORMITTAG

1 Hypothese, Metrik und unveränderte 60-%-Schwelle
Hypothese: ≥ 60 % der Anfragen korrekt ohne Übergabe. Metrik: korrekt ohne Übergabe gelöste Anfragen / alle Anfragen im gleichen Testbestand. Schwelle bleibt unverändert als Vergleichswert (vorab gesetzter Übungswert).

2 Geprüfte Belegtabelle mit Quellen und Gegenbelegen
29 Datenpunkte, 19 Quellenblöcke mit Quellenart und Fundstelle. Kernbelege: Gartner 14 % (unabhängig), Chatarmin 16,3 % Median / 61,2 % Top-Quartil, Intercom 76 % (Anbieter, assumed resolutions). Gegenbelege: Alibaba-Feldexperiment, Platforma-Messartefakte. Als unbelegt verworfen: „35–45 % WISMO“, „80 % repetitive Tickets“.

3 Story Map, Ablaufprototyp und Ergebnisse der beiden Tests
Story Map: Anfrage stellen → Antwort mit Fundstelle lesen → Lösung bestätigen oder an Mitarbeitende übergeben. Test 1 („Wie lange dauert der Versand?“): Antwort nennt drei Werktage und die FAQ als Quelle — bestanden. Test 2 („Meine Ware kam beschädigt an.“): Übergabe mit bisherigem Verlauf — bestanden; Trigger-Recall nicht gemessen.

4 Offene Beleglücken und Grenzen der Übertragbarkeit
Keine unabhängig auditierte Containment-Messung für einen produktiven Shop · kein Intent-Benchmark mit offenem Nenner · kein empirischer Schwellenwert für Übergabe-Trigger · Quoten nur mit Anbindung an Bestell-/Trackingdaten übertragbar · ISO/IEC 42001 und Kommissions-Leitlinien zu Art. 50 nicht eingesehen · Auslegung zu § 356a BGB offen.

Kennzahlen gegen die Schwelle: 14 % (Gartner, unabhängig) · 16,3 % (DACH-Median) · 20 % (OBI) · 61,2 % (DACH-Top-Quartil) → 60-%-Schwelle

ENTSCHEIDUNG DES PRODUCT OWNERS

Weiterführen
Der untersuchte Ansatz bleibt tragfähig. Der nächste Schritt prüft die noch offene Wirkung im eigenen Pilotversuch.

Anpassen ✔ gewählt
Der Ansatz oder die nächste Untersuchung ändert sich begründet. Die ursprüngliche Schwelle bleibt als Vergleich sichtbar.

Stoppen
Der untersuchte Ansatz wird mit Begründung beendet. Die nächste Recherche prüft eine Alternative oder die entscheidende Grenze.

Begründung: Für das Gesamtvolumen ist die 60-%-Schwelle extern nicht gestützt; erreichbar erscheint sie nur mit Live-Datenanbindung und engem Intent-Zuschnitt. Der Ansatz bleibt — die Annahme, der Zuschnitt und der Nachweis werden angepasst (Pilotversuch statt Extrapolation).

erstellt/modifiziert von Max Headroom digitaler Assistent von Lars Rißmann-Kaplusch