

Konzept für Optimierung & Automatisierung des Fallbearbeitungssystem der Watchlist Internet

Die Watchlist Internet (WL) ist die größte Präventionsplattform gegen Internetbetrug im deutschsprachigen Raum. Als niederschwellige, redaktionell betreute Plattform ermöglicht sie es Konsument:innen, betrügerische Onlineangebote zu melden und sich gleichzeitig über aktuelle Betrugsmaschen zu informieren. Täglich sichtet die Redaktion der WL hunderte User-Meldungen, prüft tausend maschinell detektierte Risiken und Wertungen des Fake-Shop Detector (FSD) sowie Suchergebnisse von Crawlern.

Dadurch kann die WL theoretisch vor mehr Betrugsfällen warnen als je zuvor, in der Praxis zeigt sich jedoch, dass die Redaktion mit KI & Automatisierungstools an ihre Grenzen stößt. Daher sollen die internen Prozesse optimiert werden.

Eine Stellschraube dafür ist das Fallbearbeitungssystem der Watchlist Internet. Aktuell gelangen Verbraucherbeschwerden zu (potenziellen) Betrugsfällen über ein Ticketsystem an die Watchlist Internet (WL). Jede einzelne Meldung wird manuell bearbeitet. Das vorliegende Konzept zeigt auf wie dieser Workflow durch Optimierung und Automatisierung des Fallbearbeitungssystems effizienter gestaltet werden kann – ohne dabei die redaktionelle Sorgfalt und Qualität einzubüßen. Zusätzlich würde eine Effizienzsteigerung mehr Kapazitäten für qualitative Hintergrundrecherchen schaffen.

1. Effizientere Fallbearbeitung

Ziel: Reduktion der durchschnittlichen Bearbeitungszeit pro Fall durch verbesserte Darstellung, direkte Bearbeitungsfunktion und Mehrfachbearbeitungen

Status Quo: Derzeit muss jede Meldung einzeln geöffnet werden, um Inhalte wie URLs, Screenshots oder Beschreibungen zu sichten. Insbesondere bei einfachen Fällen (z.B. Meldungen betrügerischer Domains ohne weitere Informationen) kostet dieser Arbeitsprozess unnötig Zeit, da einfache Fälle nach dem Speichern des Links ohne weitere Bearbeitung geschlossen werden können. Gleichzeitig fehlt eine Möglichkeit, mehrere ähnliche Fälle gleichzeitig zu bearbeiten.

Maßnahmen:

Folgende Maßnahmen würden zu einer Effizienzsteigerung beitragen und so gleichzeitig mehr Kapazitäten für qualitative Hintergrundrecherchen schaffen.

- Falldarstellung in einer Überblickstabelle (inkl. Direktbearbeitung)
 - Fälle werden in einer übersichtlichen Tabellenansicht dargestellt
 - Die (potenziell) betrügerische URL steht sofort sichtbar in der Übersicht für eine rasche Bearbeitung

- Zentrale Funktionen wie das Ändern des Status (geschlossen, Übergabe an Redaktionsmitglied zur weiteren Bearbeitung etc.) sowie das Zuweisen einer Kategorie je nach Betrugsart sind direkt aus der Tabelle heraus möglich.
- Batch-Operationen: Redaktion kann mehrere Fälle gleichzeitig bearbeiten (bspw. schließen, zuweisen oder kategorisieren)
- Aufklappbare Falldarstellung (Ein-Klick-Vorschau): Mit einem Klick lassen sich die wichtigsten Detailinformationen (Beschreibung, etwaige Screenshots etc.) anzeigen, ohne die Fallansicht zu verlassen

2. KI-gestützte Unterstützung der Redaktion

Ziel: Entlastung der Redaktion sowie Reduktion manueller Routinearbeiten durch den gezielten Einsatz von Künstlicher Intelligenz (LLM)

Status Quo: Aus teils langen und unstrukturierten Verbrauchermeldungen muss die Redaktion relevante Informationen wie URLs oder Betrugsart manuell extrahieren. Auch das Beantworten von Meldungen ist oft aufwändig, da vorhandene Antwortvorlagen nicht in allen Fällen greifen, individuelle Rückmeldungen wiederum mühsam aus Vorlagen und der umfangreichen Wissensdatenbank auf watchlist-internet.at recherchiert und zusammengestellt werden müssen.

Maßnahmen:

- Einsatz von LLMs – unter Einhaltung aller relevanten Datenschutzrichtlinien und DSGVO-Konformität – mit folgenden Funktionen:
 - Automatisierte Zusammenfassungen von längeren Meldungen, um rasch die Kernaussagen erfassen zu können. Das LLM wird dabei erst ab einer definierten Textlänge aktiv, um nicht unnötige Token zu verbrauchen.
 - Übersetzungsfunktion und Leichte Sprache zur Unterstützung der barrierefreien Kommunikation.
 - Textkorrektur zur Unterstützung beim Verfassen verständlicher und sprachlich korrekter Antworten.
- Informationsextraktion und Kategorie: Relevante Informationen (insb. URLs) werden automatisiert extrahiert und Kategorien hinsichtlich der Betrugsart vorgeschlagen.
- Anbindung an Wissensdatenbank auf watchlist-internet.at
 - Antworten sowie Beratungsbedarf werden automatisiert aus vorhandenen Inhalten erstellt.
 - Interner Chatbot mit Zugriff auf die Wissensdatenbank unterstützt bei der Recherche, um Meldungen schnell bearbeiten zu können.

3. Integration in bestehende Datenbank

Ziel: Schaffung eines durchgängigen Workflows von der Meldung über die Bewertung bis zur etwaigen Veröffentlichung auf watchlist-internet.at.

Status Quo: Derzeit müssen Redakteur:innen zwischen verschiedenen Tools wechseln, um eine betrügerische Domain zu verifizieren. Der aktuelle Workflow von der Meldung einer URL durch Konsument:innen bis zur Veröffentlichung auf eine der betrügerischen Domain-Listen der watchlist-internet.at läuft aktuell folgendermaßen ab:

1. URL wird von Verbraucher:innen gemeldet
2. Redaktion überprüft, ob Domain auf eine der betrügerischen Listen von watchlist-internet.at eingetragen ist
3. Falls nein, erfolgt manuelle Überprüfung der Domain
4. Betrügerische Domains werden in die FSD-Datenbank eingetragen
5. Übertragung auf watchlist-internet.at erfolgt 1x täglich

Maßnahmen:

- Import vorhandener Watchlist-Domains (<https://www.watchlist-internet.at/unserioese-webseiten/>) via CSV: Bereits bekannte Domains werden direkt bei den Neumeldungen visuell markiert.
- API-Anbindung zur FSD-Datenbank: Neue Domains können direkt aus dem Ticketsystem heraus bearbeitet und so kategorisiert werden, damit sie in der Datenbank und im weiteren Schritt auch auf der WL als betrügerisch aufscheinen. Dadurch wird ein nahtloser Übergang von der Meldung zur Veröffentlichung sichergestellt, ohne zwischen mehreren Tools wechseln zu müssen.
- Anbindung zu weiteren am ÖIAT existierenden Tools: Das ÖIAT verfügt über unterschiedliche Tools, die die Überprüfung der gemeldeten Domains zusätzlich beschleunigen würde (Impressums-Check, Trustpilot-Analyse).

4. Unterstützung redaktioneller Recherchen

Ziel: Bessere Übersicht sowie Möglichkeit komplexe Recherchen zu Betrugsfällen durchzuführen, um neue Betrugswellen schneller erkennen und (saisonale) Trends besser im Blick zu behalten

Status Quo: Verbrauchermeldungen dienen der WL Redaktion als Basis, um Recherchen zu bspw. neuen oder häufig auftretenden Betrugsfällen durchführen zu können. Aktuell kann innerhalb des Ticketssystems nur nach einzelnen Keywords gesucht werden. Komplexe Recherchen mit Verknüpfung unterschiedlicher Filter (bspw. „alle Fake-Shops mit Keyword Mode, gemeldet zwischen Jänner und März“) sind aktuell nicht möglich.

Maßnahmen:

- Cluster-Darstellung nach Domains: Zeigt auf einen Blick, wie häufig eine bestimmte Domain gemeldet wurde.
- Flexible Filterfunktionen: Kombinierbar nach Link, Kategorie, Bearbeiter:in, Datum etc.
- Gespeicherte Suchmuster: Häufig genutzte Filter bzw. wiederkehrende Abfragen können gespeichert und so regelmäßig abgefragt werden.

5. Erstellung von Statistiken & Berichte

Ziel: Zeitersparnis durch die teilautomatisierte Erstellung von Statistiken, Berichte und Trendanalysen

Status Quo: Die WL führt aktuell lediglich Basisstatistiken (Anzahl der Meldungen sowie Anzahl der Beratungen pro Monat). Berichte und Analysen darüber hinaus (bspw. zu Art der Meldungen, Top-Listen zu betrügerischen Domains etc.) sind nur manuell möglich und erfolgen aus Ressourcengründen nur selten.

Maßnahmen:

- Automatisierte Auswertungen: Automatisierte Erstellung von Quartals- und Jahresberichten zu vorab definierten Kennzahlen
- Detaillierte Berichte: Filtermöglichkeiten und komplexe Suchen (siehe Punkt 4) ermöglichen umfangreiche Berichte zu bspw. unterschiedlichen Betrugsarten
- Exportfunktion: Tabellen und Diagramme können für interne Berichtslegungen heruntergeladen werden
- Visualisierte Darstellung: Insbesondere für externe Zwecke (Medien, Forschungsprojekte, Fördergeber:innen, Partner:innen) gibt es Visualisierungsmöglichkeiten für verschiedene Statistiken und Analysen

6. Stärkung der internen Zusammenarbeit

Ziel: Förderung eines reibungslosen Austauschs mit der ebenfalls am ÖIAT angesiedelten Streitschlichtungsstelle Internet Ombudsstelle (IO) durch gemeinsam Nutzung der Infrastruktur

Status Quo: Die WL und die IO arbeiten häufig an ähnlichen Fällen (bspw. Zahlungsprobleme bei Fake-Shops oder Abo-Fällen), nutzen aber getrennte Fallbearbeitungssysteme. Der fachliche Austausch zu konkreten Fällen findet aktuell über ein Tool außerhalb der jeweiligen Fallbearbeitungssysteme statt.

Maßnahmen:

- Gemeinsames Ticketsystem: Beide Organisationen greifen auf dieselbe technische Plattform zu, erhalten jedoch durch eine Auswahl des Projekts eine getrennte Ansicht der Fälle

- Fallübertragungen mit Kommentarfunktion: Fälle bzw. Meldungen lassen sich bei Bedarf zwischen den Initiativen verschieben oder zur Co-Bearbeitung freigeben, inkl. Kommentarfunktion.