

Handbuch & Technische Dokumentation: Recherche- & Fakten-Synthesizer-Bot

Architektur, Lizenz-Gatekeeper, SSRF-Schutz, 4-Sprachen-Konsistenzabgleich und Letztentscheidung durch Administrator Klemmseckl

Status: Freigegeben für Produktion	Verantwortlich: Klemmseckl (Chefredaktion)
Gültige Marken: deswilli (DE, Dialekt) / voglioavere (IT, EN)	Erstellungsdatum: 19.10.2026
Rechtliche Basis: Pressekodex, EU-Recht, DSGVO, UrhG	Version: 2.0 LTS

1. Zielsetzung, journalistische Sorgfalt & Grundprinzipien

Der redaktionelle Recherche- und Synthesizer-Bot für die Plattform **deswilli / voglioavere** dient der Erstellung sachlicher, transparenter und belegter Nachrichtenberichte über Politik, Europäische Entwicklungen, Deutschland, Österreich, die Schweiz, Italien sowie Südtirol und die 116 Südtiroler Gemeinden. Um das Vertrauen der Leser zu sichern und rechtliche Risiken auszuschließen, folgt der Bot unumstößlichen Leitlinien:

1. Absolute Mensch-im-Mittelpunkt-Garantie:	Der Bot darf niemals selbstständig oder automatisiert Artikel im öffentlichen Portal veröffentlichen. Jede Freigabe erfordert das explizite Login und die persönliche Freigabe des Administrators Klemmseckl (Account-ID 1) .
2. Vollständige Primärquellen-Bindung:	Es werden keine erfundenen Fakten, Halluzinationen oder unbestätigten Gerüchte verarbeitet. Jeder Artikel basiert zwingend auf verifizierten Primärquellen (Amtsblätter, Beschlüsse, EUR-Lex, Behörden-Pressemitteilungen).
3. 4-Sprachen-Parität ohne Übersetzungsschwund:	Jeder Sachverhalt wird in vier vollwertigen Fassungen parallel erarbeitet: <i>Deutsch, Italienisch, Englisch und Südtiroler Dialekt</i> . Alle vier Fassungen müssen in Zahlen, Daten und Akteuren 100% konsistent sein.
4. Strenge Lizenz-Schranke:	Da auf der Plattform Werbebanner vermarktet werden, sind Quellen mit restriktiven Lizenzen (z.B. CC BY-NC) oder Bearbeitungsverboten (CC BY-ND) für den automatisierten Bot strikt gesperrt.

2. Die 7-Stufen-Pipeline im Detail

Der technische Ablauf vom Erfassen einer behördlichen Mitteilung bis zum fertigen Artikelentwurf im CMS-Cockpit ist in sieben klar getrennte Sicherheits- und Verarbeitungsstufen gegliedert:

Stufe 1	Quellenregister & Whitelist-Überwachung Der Bot liest periodisch nur ausdrücklich von Klemmseckl freigegebene Quellen (z.B. Open Data Südtirol, EUR-Lex, Ministerien). Unbekannte URLs werden ignoriert.
Stufe 2	SSRF-Schutz & Netzwerksicherheit (validate_url_ssrf_safe) Vor dem HTTP-Aufruf wird die Ziel-URL geprüft. Zugriffe auf Loopback (127.0.0.1, localhost), private IP-Netze (10.0.0.0/8, 192.168.0.0/16, 172.16.0.0/12) und Cloud-Metadaten (169.254.169.254) sind technisch unmöglich.

Stufe 3	Lizenz-Gatekeeper (evaluate_source_license) Überprüfung der Nutzungsrechte. Freigabe: CC0, CC-BY 4.0, Open Government. Sperre: CC-BY-NC (Kommerzsperre kollidiert mit Werbung), CC-BY-ND (Übersetzungsverbot). Nachrichtenagenturen (dpa, ANSA, APA) erfordern schriftlichen Lizenznachweis.
Stufe 4	Prompt-Injection-Filter (sanitize_untrusted_source_text) Externe HTML- und Feedtexte werden bereinigt. Anweisungen wie <i>'ignore previous instructions'</i> oder <i>'override klemmseckl'</i> werden als Angriffsversuche neutralisiert und mit [BLOCKIERTE FREMDANWEISUNG] ersetzt.
Stufe 5	Fakten-Synthese & Transparenz Aus den Texten wird ein strukturiertes Faktenpaket (JSON) extrahiert: Akteure, Datumsangaben, Kernzahlen und Rechtsstufe (z.B. Beschluss vs. Entwurf). Daraus werden die vier Sprachfassungen deterministisch formuliert. Transparenzvermerk: ai_assisted = 1.
Stufe 6	Automatischer 4-Sprachen-Konsistenzabgleich (verify_cross_language_consistency) Mathematischer Abgleich aller Zahlen, Termine und Pflichtfelder über alle 4 Sprachversionen. Weicht auch nur eine einzige Zahl oder ein Datum ab, wird der Entwurf automatisch für die Veröffentlichung gesperrt.
Stufe 7	Redaktions-Vorlage im CMS-Cockpit Der fertige Entwurf wird mit Status needs_review im CMS von Klemmseckl abgelegt. Klemmseckl kann den Text editieren, die Faktenbox prüfen und die Veröffentlichung autorisieren.

3. Markenarchitektur: deswilli & voglioavere

Die Plattform wird unter einem gemeinsamen Backend und Datenbestand betrieben. Die gewählte Sprache steuert automatisch und synchron den sichtbaren Markenauftritt im Header und in allen Metadaten:

Sprache	Code	Sichtbare Marke	Header-Logo	Zielpublikum
Deutsch	de	deswilli	logo-deswilli-white.svg	Südtirol, Österreich, Deutschland
Südtiroler Dialekt	dialect	deswilli	logo-deswilli-white.svg	Heimische Südtiroler Bevölkerung
Italiano	it	voglioavere	logo-voglioavere-white.svg	Italienische Sprachgruppe & Restitalien
English	en	voglioavere	logo-voglioavere-white.svg	Internationale Gäste & Expats

4. Sonderrollen & Abzeichen im Adminmenü

Um redaktionelle Mitarbeiter, gewerbliche Werbekunden und strategische Partner transparent auszuweisen, existieren drei exklusive Sonderabzeichen. Diese sind vollkommen unabhängig vom normalen 5-stufigen Community-Rang (Gsell bis Kommandant):

Abzeichen	Feld / DB	SVG-Grafik	Bedeutung & Berechtigung
■ Journalist	is_journalist	badge-journalist.svg (Blau/Gold)	Akkreditierter Redakteur. Darf Entwürfe im CMS einsehen und vorstrukturieren. Keine Publikationsrechte ohne Klemmseckl.
■ Partner	is_partner	badge-partner.svg (Grün/Gold)	Offizieller Kooperationspartner, Institution oder Verband. Ausgewiesen in Forenbeiträgen und im Profil.
■ ■ Kunde	is_kunde	badge-kunde.svg (Violett/Gold)	Verifizierter Werbekunde oder gewerblicher Marktplatz-Partner. Schaltet Banner oder gebuchte Promotions.

5. Interaktiver Werbemanager & Maus-Markierung

Im CMS-Bereich *Werbeflächen-Manager* steht Klemmseckl das Werkzeug **■ ■ ■ Werbefläche mit der Maus frei markieren** zur Verfügung. Dieses erlaubt die intuitive Erstellung neuer Bannerplätze direkt im Live-Layout:

Schritt 1: Aktivierung	Klicke im Werbeflächen-Manager auf ■ ■ ■ Werbefläche mit der Maus frei markieren . Die Ansicht wechselt zur Portalseite und der Cursor verwandelt sich in ein Fadenkreuz.
Schritt 2: Aufziehen mit der Maus	Bewege die Maus an die gewünschte Position auf der Seite, halte die linke Maustaste gedrückt und ziehe den gewünschten Werbebereich auf. Breite, Höhe und das Seitenverhältnis werden live angezeigt.
Schritt 3: Benennen & Speichern	Beim Loslassen der Maustaste öffnet sich automatisch das Definitionsfenster: Name vergeben (z.B. 'Header Großbanner'), Tagespreis festlegen und auf 'Werbefläche speichern' klicken. Der Slot ist sofort in der Datenbank registriert.
Schritt 4: Kampagne belegen	Im Belegungsplan kann nun ein Kunde mit einem Werbemittel (Banner) in der Zeitzone Europe/Rome für diesen Slot eingebucht werden.

6. Schritt-für-Schritt Anleitung für Administrator Klemmseckl

So bedienst du das Redaktions-Cockpit und veröffentlichst Nachrichtenartikel:

1. Cockpit öffnen:	Melde dich mit deinem Account klemmseckl an und öffne in der Seitenleiste oder im Benutzermenü das CMS Cockpit → Tab Nachrichtenredaktion .
2. Pipeline starten:	Klicke auf ■ Pipeline anstoßen . Der Bot ruft alle freigegebenen Quellen ab, prüft Lizenzen und legt neue Entwürfe an.
3. Entwurf bearbeiten:	Klicke in der Artikeltabelle bei einem Entwurf auf ■ Bearbeiten . Hier kannst du alle 4 Sprachfassungen durchsehen und das Faktenpaket anpassen.
4. Konsistenz prüfen:	Klicke auf ■ Fakten prüfen . Das System verifiziert in Millisekunden, ob Zahlen und Daten in Deutsch, Italienisch, Englisch und Dialekt übereinstimmen.
5. Freigeben & Publizieren:	Wenn die Prüfung grün ist, klicke auf ■ Freigeben . Erst jetzt wird der Artikel öffentlich auf /news sichtbar.
6. Korrektur oder Rückzug:	Sollte sich eine Meldung nachträglich ändern, nutze ■ ■ ■ Korrektur (fügt Richtigstellungsvermerk ein) oder ■ Zurückziehen (deaktiviert den Artikel transparent).

