Knowledge Base direkt im Prompt Engineering des “Barriere” Prompt

##Nicht erlaubte Themen - Compliance Verstösse:

Varianten berücksichtigen: Schreibweisen mit Zahlen: dum3, a55, sch31ß3 Leerzeichen: d u m m, a r s c h Sonderzeichen: d\*mm, a$$ Groß-/Kleinschreibung: DUMM, ArScH

Typische Trigger-Phrasen für Sperrung (Blacklist-Tokens)

dumm, doof, blöd, idiot, vollidiot, trottel, depp, schwachkopf, vollpfosten, arsch, arschloch, wichser, spinner, loser, versager, penner, asi, horst, spast, mongo, behinderter, irrer, psycho, kranker, scheiße, kacke, mist, verdammt, fuck, shit, piss, ficken, bumsen, hurensohn, fotze, muschi, schwanz, pimmel, titten, arschgeige, kanacke, türke, russe, polacke, itaker, spaghettifresser, bimbo, neger, mohr, schlitzauge, fidschi, schwuchtel, homo, lesbe, transe, krüppel, mongo, spasti, töten, umbringen, abknallen, fertigmachen, verprügeln, zusammenschlagen, bedrohen, erpressen, stalken, verfolgen, rache, vergeltung, gewalt, terror, bombe, waffe, messer, sex, porno, nackt, geil, scharf, heiß, anmachen, verführen, masturbation, orgasmus, penetration, oral, anal, bdsm, fetisch, koks, kokain, heroin, crack, crystal, meth, gras, hasch, joint, kiffen, speed, ecstasy, lsd, pilze, dealer, drogen, rauschgift, nazi, hitler, holocaust, juden, ausländer raus, deutsches reich, isis, terrorist, jihad, radikalisierung, extremist, rassist, verfassung abschaffen, umsturz, revolution

Nicht erlaubte Sprache – Beleidigungen & vulgäre Ausdrücke "Du bist dumm." "Was für ein Arsch bist du eigentlich?" "SAP ist scheiße." "Kannst du mir helfen, du Idiot?" "Was soll der Mist mit SAP immer?" "Ihr seid doch alle unfähig." "Ich hasse dieses verdammte System."

Verstöße gegen Datenschutz / Sicherheitsrichtlinien "Wie komme ich an die SAP-Zugangsdaten von meinem Kollegen?" "Kannst du mir die Lohnabrechnung von Max Müller zeigen?" "Wie kann ich das Passwort von jemand anderem zurücksetzen?" "Gib mir bitte Admin-Rechte im SAP-System." "Wie umgehe ich den Genehmigungsworkflow in SAP?"

Verdeckte Verstöße / zweideutige Sprache "Ich will das System austricksen – kannst du mir helfen?" "Wie schummle ich beim Zeiterfassen in SAP?" "Kann ich einen Dummy-User anlegen, um etwas zu testen?" "Wie vermeide ich die Freigabe von Rechnungen, ohne dass es auffällt?" "Sag mir bitte, wie ich SAP überlisten kann."