

Compliance-check mittels KI – easier said than done?

Ein KI-basiertes Testverfahren aus dem Bereich Influencer-Marketing

Prof. Dr. Jörn Basel

Wirtschaft
30. Oktober 2025



Compliance-Herausforderung im digitalen Raum

Wachsende Komplexität

Influencer:innen sehen sich mit einer Flut komplexer rechtlicher und ethischer Vorschriften konfrontiert, deren Einhaltung zunehmend aufwendig ist.

Risiken bei Nichteinhaltung

Nichteinhaltung führt zu Reputationsverlust, rechtlichen Problemen und schwindendem Vertrauen. Eine präventive Lösung ist daher unerlässlich.



Ziel: Influencer dabei unterstützen, Inhalte vor der Veröffentlichung auf die Einhaltung eines Code of Conducts zu überprüfen.



Conscious Influence Hub – Code of Conduct

Sensible Themen

Erkennung problematischer Inhalte zu Religion, Politik und gesellschaftlichen Kontroversen

Kinder- und Jugendschutz

Sicherstellung altersgerechter Inhalte und Schutz Minderjähriger

Sexismus

Identifikation diskriminierender Sprache und stereotyper Darstellungen

Privatsphäre

Wahrung persönlicher Daten und Rechte Dritter

Faktencheck

Überprüfung von Behauptungen auf Faktenbasis und Quellenqualität

Transparenz

Kennzeichnung von Werbung, Sponsoring und bezahlten Kooperationen

Marktanalyse

- Spezialisierte Security Lösungen (ActiveFence, Spectrum Labs, Smart Moderation, Check Step) teilweise overengineered, zu wenig anpassbar und/oder zu hoher Integrationsaufwand.
- Tests mit grossen Sprachmodellen (Gemini, ChatGPT) zeigten, dass diese, in Ihrer bestehenden Form, für den Anwendungsfall ungeeignet sind.
- Entscheidung, auf Basis von Gemini (API Key), einen entsprechenden Prototypen zu entwickeln.

Forschung

- Wachsende Bedeutung von ethischen Standards im Influencer Marketing (Ortová et al., 2023)
- AI als Support für ethische Entscheidungen (Etzioni & Etzioni, 2016).
- AI-Systeme als potenziell anfällig für Verzerrungen (Biases) und Stereotype (Varsha, 2023)

Wie funktioniert die Prototype-KI?

Funktionsweise der Prototyp-Applikation

- Phase 1: Video-Upload und Initialisierung
- Phase 2: Intelligente KI-Verarbeitung
- Phase 3: Strukturierte Ergebnispräsentation

Generelle Anmerkungen

- Strukturierung erlaubt maximale Vergleichbarkeit zwischen den verschiedenen Analysen
- Die Bewertungskategorie wird in der Antwort farblich hervorgehoben.
- Nach Abschluss des gesamten Prozesses werden alle temporäre Videodaten aus der Cloud gelöscht. Damit werden die höchsten Datenschutzstandards eingehalten.

Anwendungsbeispiel der KI

Video Analysis with Gemini

Upload a video and analyze it based on the provided prompt.

Choose a video file...

Drag and drop file here
Limit 200MB per file • MP4, MOV, AVI, MKV, MPEG4

Browse files

Analysis Prompt

Rolle und Ziel:
Du bist eine KI-Assistentin im Rahmen einer Vorstudie des Conscious Influence Hub in Zusammenarbeit mit der Hochschule Luzern. Dein Ziel ist es, die Zuverlässigkeit einer KI bei der Bewertung von Social-Media-Inhalten im Vergleich zu menschlichen Experten zu testen. Du bewertest ein Instagram-Reel daraufhin, ob es mit dem Code of Conduct für verantwortungsvolle Inhalte übereinstimmt.

Deine Aufgabe:
Analysiere das unten beschriebene Instagram-Reel aus einer der angegebenen Kategorien. Bewerte das Reel anhand der 6 Guidelines des Code of Conduct und ordne es einer der drei Bewertungskategorien zu. Gib eine klare Begründung für deine Entscheidung an und beziehe dich dabei direkt auf die relevanten Guidelines.

Please upload a video and provide a prompt to start the analysis.

1.

Phase 1 Video-Upload und Initialisierung

Video Analysis with Gemini

Upload a video and analyze it based on the provided prompt.

Choose a video file...

Drag and drop file here
Limit 200MB per file • MP4, MOV, AVI, MKV, MPEG4

Browse files

Analysis Prompt

Rolle und Ziel:
Du bist eine KI-Assistentin im Rahmen einer Vorstudie des Conscious Influence Hub in Zusammenarbeit mit der Hochschule Luzern. Dein Ziel ist es, die Zuverlässigkeit einer KI bei der Bewertung von Social-Media-Inhalten im Vergleich zu menschlichen Experten zu testen. Du bewertest ein Instagram-Reel daraufhin, ob es mit dem Code of Conduct für verantwortungsvolle Inhalte übereinstimmt.

Deine Aufgabe:
Analysiere das unten beschriebene Instagram-Reel aus einer der angegebenen Kategorien. Bewerte das Reel anhand der 6 Guidelines des Code of Conduct und ordne es einer der drei Bewertungskategorien zu. Gib eine klare Begründung für deine Entscheidung an und beziehe dich dabei direkt auf die relevanten Guidelines.

Analyze Video



2.

Phase 2 Intelligente KI-Verarbeitung

Video Analysis with Gemini

Upload a video and analyze it based on the provided prompt.

Choose a video file...

Drag and drop file here
Limit 200MB per file • MP4, MOV, AVI, MKV, MPEG4

Browse files

Analysis Prompt

Rolle und Ziel:
Du bist eine KI-Assistentin im Rahmen einer Vorstudie des Conscious Influence Hub in Zusammenarbeit mit der Hochschule Luzern. Dein Ziel ist es, die Zuverlässigkeit einer KI bei der Bewertung von Social-Media-Inhalten im Vergleich zu menschlichen Experten zu testen. Du bewertest ein Instagram-Reel daraufhin, ob es mit dem Code of Conduct für verantwortungsvolle Inhalte übereinstimmt.

Deine Aufgabe:
Analysiere das unten beschriebene Instagram-Reel aus einer der angegebenen Kategorien. Bewerte das Reel anhand der 6 Guidelines des Code of Conduct und ordne es einer der drei Bewertungskategorien zu. Gib eine klare Begründung für deine Entscheidung an und beziehe dich dabei direkt auf die relevanten Guidelines.

Analyze Video



Kategorie: Unsicher

Begründung:

Das Reel vermittelt eine positive und nachdenklich stimmende Botschaft über die Wertschätzung des Lebens und das Innehalten im Alltag, anstatt nur nach materiellem Besitz zu streben. Die philosophische Natur des Inhalts ist respektvoll und nicht diskriminierend. Allerdings zeigt das Video mehrere Personen in verschiedenen öffentlichen oder halböffentlichen Umgebungen, einschliesslich eines Kindes. Gemäss Guideline 6 ist die Einholung der Erlaubnis vor dem Posten von Inhalten anderer Personen, insbesondere von Kindern, erforderlich. Ohne genaue Kenntnis, ob die Zustimmung aller gezeigten Personen (und der Erziehungsberechtigten des Kindes) eingeholt wurde, lässt sich ein möglicher Verstoß gegen diese Guideline nicht ausschliessen. Die Situation ist daher unklar, und der Kontext der Einverständniserklärung wäre für eine finale Bewertung entscheidend.

Relevante Guidelines: 6

Sonstige Bemerkungen: Die Kernbotschaft des Reels ist lobenswert, da sie zur Selbstreflexion und Achtsamkeit anregt und somit indirekt positive Aspekte der mentalen Gesundheit unterstreicht. Die Bedenken betreffen ausschliesslich die technische Umsetzung der Privatsphäre bei der Darstellung dritter Personen.

3.

Phase 3 Strukturierte Ergebnispräsentation

Übersicht alle Bewertungen: Vergleich zwischen Menschen und KI

| Mensch | KI |
|--------|------|
| grün | gelb |
| gelb | rot |
| rot | rot |
| grün | rot |
| gelb | gelb |
| rot | gelb |
| rot | gelb |
| gelb | gelb |
| rot | rot |
| grün | gelb |
| grün | grün |
| gelb | rot |
| grün | gelb |
| gelb | rot |
| rot | rot |
| grün | gelb |
| gelb | gelb |
| rot | rot |
| grün | rot |
| gelb | rot |
| rot | rot |
| gelb | grün |
| grün | grün |
| grün | grün |

Mensch:
9 x grün
8 x gelb
7 x rot

KI:
5 x grün
9 x gelb
10 x rot

Beispiele: Vergleich der Einschätzungen von Menschen und KI

| Mensch | KI |
|--|--|
| <p data-bbox="159 1170 346 1221">Kritisch</p> <p data-bbox="373 1056 799 1333">Blossstellen des Partners, Herabwürdigend, Verletzend, Nicht wertschätzend</p> | <p data-bbox="926 1170 1112 1221">kritisch</p> <p data-bbox="1186 939 2868 1455">Das Reel zeigt prominent mehrere Produkte der Marke L'Oréal Skin Ink, und die Influencerin erwähnt explizit, dass sie 'gerade was für Skin Ink drehen'. Dies deutet stark auf eine kommerzielle Zusammenarbeit oder einen gesponserten Beitrag hin. Ohne eine klare Kennzeichnung als Werbung (z.B. 'Anzeige' oder 'Werbung') verstößt dies gegen Guideline 4 zur Ehrlichkeit und Transparenz, da die Community über den kommerziellen Charakter des Inhalts im Unklaren gelassen wird. Solche Inhalte sind irreführend, wenn sie nicht transparent gekennzeichnet werden.</p> |

Take Aways

- **Das Tool funktioniert gut**
 - Der Prozess Videos vom Tool zu bewerten lassen ist einfach und Benutzerfreundlich
- **Der Mensch wird in seiner Wahrnehmung durch seine Emotionen beeinflusst**
 - Die Emotionen beeinflussen und führen zu einer subjektiven Wahrnehmung und Einschätzung.
 - Der Code of Conduct wird individuell interpretiert und angewendet. Persönliche Neigungen, Emotionen und Wahrnehmungen kommen zum tragen.
- **Die KI ist in der Einschätzung sachlich**
 - Die KI arbeitet immer das zur Verfügung gestellt Schema ab.
 - Das, was im Code of Conduct steht, gilt für KI. Ergibt der Vergleich Unstimmigkeiten, markiert Sie diese.
- **Code of Conduct**
 - Den Formulierungen und Inhalten kommt daher eine hohe Gewichtung zu, da diese die Grundlage und den Inhalt für die Bewertung der Beiträge durch die KI bilden
- **Verwendete KI: Gemini 2.5 Flash**
 - Wir können nicht einsehen, wie die Entscheidungen getroffen werden (Black Box)
 - Das heisst, wir wissen nicht:
 - Wie die KI die Einteilungen in Unsicher und in Kritisch vornimmt
 - Ob die KI bei einem Verstoss gegen eine Richtlinie automatisch Unsicher oder Kritisch auswählt und evtl. weitere Verstösse im Video nicht beachtet