

Die Metadaten Lounge in der IG Produktmetadaten

Juni 2024 – KI Opt-out und KI-generierten Content kennzeichnen

Präsentiert von Herbert Barkmann und Carsten Wehmeyer

**Börsenverein des
Deutschen Buchhandels**
IG Produktmetadaten



KI Opt-out: Ausgangssituation

- Urheberrechtlich geschützte (digitale) Inhalte wurden zum Trainieren von KI-Systemen verwendet
 - Cover von Büchern
 - E-Book-Inhalte
 - Stimmen von Sprecher*innen



ChatGPT vollendet Game-of-Thrones-Saga

ChatGPT hat geschafft, was Autor George R.R. Martin bisher nicht vermochte: Die KI hat die Buchreihe Das Lied von Eis und Feuer komplettiert. Allerdings überleben die wichtigen Charaktere.

[in Pocket speichern](#) [merken](#)

21. Juli 2023, 8:45 Uhr, Rainer Schneider



Autor George R.R. Martin hat Game of Thrones selbst nicht fertiggeschrieben (hier bei der Emmy-Verleihung 2019).

<https://www.golem.de/news/ki-chatgpt-vollendet-game-of-thrones-saga-2307-176040.html>

You
Gib mir eine Zusammenfassung von Eragon Teil 1

ChatGPT
Selbstverständlich, hier ist eine Zusammenfassung von "Eragon", dem ersten Teil der Inheritance Cycle von Christopher Paolini:

"Eragon" erzählt die Geschichte von Eragon, einem Bauernjungen aus dem abgelegenen Dorf Carvahall. Eines Tages entdeckt er einen glänzenden blauen Stein, der sich als Drachenei herausstellt. Der Drache, den Eragon aus diesem Ei schlüpfen lässt, wird von ihm Saphira genannt.

Als das Dorf von den Schergen des tyrannischen Königs Galbatorix angegriffen wird, muss Eragon fliehen. Gemeinsam mit seinem Mentor Brom macht er sich auf eine gefährliche Reise. Brom, ein ehemaliger Drachenreiter, unterweist Eragon in den Künsten der Magie und des Schwertkampfes.

Die Geschichte entwickelt sich zu einem epischen Abenteuer, als Eragon und Saphira auf die Spur der Varden stoßen, einer Widerstandsgruppe gegen Galbatorix. Unterwegs erfährt Eragon von seiner eigenen Bestimmung als Drachenreiter und dem Erbe einer uralten Zivilisation.

Der Konflikt zwischen Eragon und Galbatorix spitzt sich zu, während Eragon seine magischen Fähigkeiten entwickelt und eine Rebellion gegen den brutalen König unterstützt. Das Buch endet mit einem dramatischen Showdown, der den Weg für die Fortsetzung der Geschichte ebnet.

You
Was passiert in Kapitel 2?

ChatGPT
Leider kann ich nicht spezifisch auf einzelne Kapitel von Büchern eingehen, da dies zu detailliert und umfangreich ist, und ich keinen Zugriff auf Kapitel-weise Zusammenfassungen von urheberrechtlich geschützten Werken habe.

Wenn du spezifische Fragen zu Handlungselementen, Charakteren oder Themen in "Eragon" hast, die nicht zu detailliert sind, kann ich mein Bestes tun, um dir zu helfen. Du könntest mir zum Beispiel eine spezifischere Frage stellen oder allgemein nach bestimmten Aspekten des Buches erkundigen.

You
Gib mir eine Zusammenfassung Eragon Teil 2

ChatGPT
Natürlich! Im zweiten Teil von Christopher Paolinis Eragon-Saga, namens "Eragon – Der Auftrag des Ältesten", setzt der junge Drachenreiter seine epische Reise fort. Hier ist eine Zusammenfassung:

Nachdem Eragon den ersten Schritt auf seinem Weg als Drachenreiter gemacht hat, begibt er sich auf die Suche nach dem einzigen verbliebenen Mitglied des Drachenordens, den Ältesten. Zusammen mit seinem Drachen Saphira und seinem Begleiter Brom, der ihn in den Wegen der Magie unterweist, reist er durch das Reich Alagaësia.

Während ihrer Reise begegnen sie zahlreichen Gefahren, darunter den gefährlichsten Riesen, Drogen des



KI Opt-out: Problemdarstellung

- Fehlende Sensibilität für Urheberschutz bei KI-Systembetreibern
- Fehlende maschinenlesbare Kennzeichen für Urheberschutz
- Kein (etablierter) Standard für Urheberschutzkennzeichen vorhanden
- Folgen
 - Klagen von Rechteinhaber*innen
 - Offene Briefe an die Regierungen (DE, US)
 - **Ruf nach Standards und rechtlichen Vorgaben**

"Wir stellen uns keineswegs insgesamt der Entwicklung generativer KI entgegen", betont Kraus vom Cleff. "Doch die Ausbeutung unserer Werke und Leistungen durch KI-Anbieter, die wirtschaftlich von der Nutzung der Werke profitieren aber keinen Beitrag zur Refinanzierung der Kulturgüter leisten, ist nicht hinnehmbar." 27.11.2023

Offener Brief des Netzwerks Autorenrechte 21.11.2023:

*Sehr geehrter Bundeskanzler Olaf Scholz,
Sehr geehrter Bundesminister für Wirtschaft und Klimaschutz Robert Habeck,
Sehr geehrter Bundesminister für Digitales und Verkehr Volker Wissing,
sehr geehrte Repräsentanten Deutschlands in der EU:*

Mit großer Bestürzung nehmen wir, die Mitglieder des Netzwerk Autorenrechte, das sämtliche Autoren und Übersetzer des Buchsektors aus 15 Verbänden der D-A-CH-Region repräsentiert, die neue Haltung Deutschlands zur geplanten KI-Grundverordnung zur Kenntnis, die den zuvor erreichten Konsens der EU-Mitgliedsstaaten zur gesetzlichen Regulierung von KI und insbesondere der Auskunftspflichten und Risikoverantwortung für Entwickler generativer Informatik unterläuft.

Laut den Berichten von Euractiv am 19.11.2023 will Deutschland unter der Meinungsführung des Digitalministeriums gemeinsam mit dem Bundesministerium für Wirtschaft und Klimaschutz, und zusammen mit Frankreich und Italien statt einer gesetzlich verpflichtenden Regulierung auf „obligatorische Selbstregulierung“ drängen. Sanktionen für Sicherheitsverletzungen, wie etwa Urheberrechts- und Datenschutzverletzungen, mangelnde Kennzeichnung und Unterlaufen ethischer Standards, sind in dieser verstörend kultur- und bürgerfeindlichen Haltung der drei Länder nicht vorgesehen.

Herr Bundeskanzler, Herr Vizekanzler, Herr Bundesminister:

Wir appellieren an Sie, diese Haltung zugunsten vermeintlicher Wirtschaftsvorteile und zuungunsten nachhaltiger gesetzlicher Regeln aufzugeben. Ihre Haltung transportiert ein fatales Signal an alle Kulturschaffenden, an alle Bürgerinnen und Bürger: Dass Sie bereit sind, ausgerechnet jene technischen Unternehmen zu schützen, die sich für die Entwicklung ihrer Profite von Kulturwerken und Bürgerdaten illegitim bedienen – anstatt jene, ohne deren Arbeitsleistung und Privatdaten sämtliche Basismodelle und generativen Anwendungen gar nicht existierten.

Die Folgen Ihrer Haltung wären verheerend. Generative Informatik bedroht bereits jetzt zahlreiche Arbeitsplätze. Bereits jetzt haben sich zahlreiche schädigende „Geschäftsmodelle“ durch KI-Produkte, die sich in den Markt drängen, und eine Zunahme von Desinformation etabliert. Es wurde nachgewiesen, dass die Grundlagen für generative KI aus nicht rechtmäßig erworbenen Werken bestehen, deren Urheber dieser Nutzung weder zugestimmt haben noch darüber informiert wurden. Ohne gesetzliche Regelung beschleunigen generative Technologien die Ausweitung von Werks- und Datendiebstahl, sie vertiefen Diskriminierung, Informations-Verfälschung inklusive Reputationsschädigung, und leisten massiv der Klimaschädigung Vorschub. Je mehr generative Produkte gesetzlich dereguliert den Markt erreichen, desto irreparabler wird der gesamtgesellschaftliche Vertrauensverlust in Texte, Bilder und Informationen.

Wir appellieren an Sie, sich auf die Werte von Vertrauen, Demokratie und Gerechtigkeit zu besinnen. Wir stehen an der Schwelle einer Evolution, an einem der entscheidendsten Momente der künftigen Geschichte. Regulieren wir Maschinen, die sich von Menschen bedienen, um Menschen zu ersetzen – oder wählen wir die kurzsichtige Ideologie des Geldes.

Wir hoffen auf Ihr politisches Rückgrat, das Richtige zu tun.

<https://www.boersenblatt.net/news/netzwerk-autorenrechte-fordert-gesetzliche-ki-regulierung-310441>



KI Opt-out: Rechtliche Informationen

- Basis: EU AI Act / KI-Verordnung / § 44b UrhG
 - Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in **maschinenlesbarer** Form erfolgt.
 - TDM für Wissenschaft und Forschung ist grundsätzlich erlaubt
 - Transparenz: KI-Unternehmen müssen eine „**hinreichend detaillierte Zusammenfassung**“ der für das Training verwendeten Inhalte nach einer Vorlage des AI Office erstellen (ISBN, Quelle, Metdaten)



KI Opt-out: technische Basis

- Das Text & Data Mining Reservation Protocol wird genutzt, um Ausprägungen der Verarbeitungsarten anzubieten.
- Das Attribut heißt **tdm-reservation**
- Das Attribut kann 2 verschiedene Werte haben:
 - **1**: Die Verarbeitung für Text- und Datamining ist eingeschränkt (=Rechte sind reserviert).
 - **0**: Verarbeitung für Text- und Datamining ist ohne Einschränkungen erlaubt (=Rechte sind nicht reserviert).
- Es kann (bei 1) ein **Verweis auf eine Richtlinie** hinterlegt werden.
 - Die Ausprägung beinhaltet eine URL, die den textlichen Inhalt der Richtlinie als .json zurückgibt.

<https://w3c.github.io/tdm-reservation-protocol/spec/>

<https://w3c.github.io/tdm-reservation-protocol/spec/#sec-policy>



KI Opt-out: technische Umsetzung

- Text im **Impressum**
- Schutz von Webseiten in tdmrep.json
- Schutz von **Covern im JPG**
- Schutz im **EPUB**

Der Verlag behält sich das Text- und Data-Mining nach § 44b UrhG vor, was hiermit Dritten ohne Zustimmung des Verlages untersagt ist.

```
{  
  "location": "/",  
  "tdm-reservation": 1  
}
```

<http://ns.useplus.org/ldf/vocab/DMI-PROHIBITED-AIMLTRAININ>

Siehe <https://www.iptc.org/news/exclude-images-from-generative-ai-iptc-photo-metadata-standard-2023-1/>
<https://iptc.org/news/iptc-publishes-metadata-guidance-for-ai-generated-synthetic-media/>

```
<package version="3.0" unique-identifier="BookId" prefix="tdm:  
https://boersenverein.de/tdm" xmlns="http://www.idpf.org/2007/opf">  
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/"  
    xmlns:opf="http://www.idpf.org/2007/opf">  
    <dc:identifier id="BookId">urn:uuid:3d5ea832-4a80-40e9-acd1-  
8cf57031408a</dc:identifier>  
    <dc:language>de-DE</dc:language>  
    <dc:title>TDM im EPUB</dc:title>  
    <meta property="tdm:reservation">1</meta>  
    <meta property="tdm:policy">https://www.musterverlag.de/policies/tdm-  
policy.json</meta>  
  </metadata>
```



KI Nutzung ausschließen

- Achtung: Basisfunktionen sind in ONIX 3.0 vorhanden, erst in 3.1.1 werden relevante Erweiterungen angeboten!
- ONIX 2.1 ist damit endgültig out!
- Neuigkeiten in ONIX 3.0:
 - EpubUsageConstraint für CONTENT!
- Neuigkeiten in ONIX 3.1:
 - Hybrid Open Access (Kapitel / Datum)
 - SequenceNumber in TextContent und SupportingResource
 - entfernt : AudienceCode, Conference, CurrencyZone, DateFormat, PromotionContact, Reissue
- Neuigkeiten in ONIX 3.1.1:
 - KI TDM Opt Out für TextContent und SupportingResource
 - Copyright Statement Text
 - Prize Identifier



KI Nutzung ausschließen

- Opt Out kann und sollte auf mehreren Ebenen erfolgen:
 - Inhalt des Werkes per Metadaten (ONIX)
 - Inhalt des Werkes in der Datei selbst (epub, PDF, html)
 - Inhalt von Assets (Leseproben, Cover, Illustrationen usw.) per Metadaten
 - Inhalt von Assets in der Datei selbst (epub, PDF, jpg, png usw.)
- Die ONIX Erweiterung bezieht sich nur auf die Kennzeichnung in den Metadaten der ONIX Meldungen, nicht auf die Metadaten INNERHALB der einzelnen Dateien.



KI Nutzung ausschließen - Content

- Der eigentliche Content eines Werkes ist urheberrechtlich geschützt. Die Nutzung für TDM konnte schon bisher durch EpubUsageConstraint ausgeschlossen werden:

```
<EpubUsageConstraint>
  <EpubUsageType>11</EpubUsageType>      <!-- TDM, in Liste 145 -->
  <EpubUsageStatus>03</EpubUsageStatus> <!-- is prohibited, Liste 146 -->
</EpubUsageConstraint>
```

- Zusätzlich kann eine “Lizenz” mitgegeben werden, die das Recht für TDM genauer regelt oder eine Lizenz dafür anbietet:

```
<EpubLicense>
  <EpubLicenseName>MyPublisher TDM License Policy</EpubLicenseName>
  <EpubLicenseExpression>
    <EpubLicenseExpressionType>20</EpubLicenseExpressionType> <!-- ODRL license, list 218 -->
    <EpubLicenseExpressionLink>https://www.mypublisher.com/.policies/contentpolicy.json
  </EpubLicenseExpression>
</EpubLicense>
```

- Der Wert 20 ist NEU in Liste 218 in Version 65 der ONIX Codelisten!



KI Nutzung ausschließen - Lizenz

- ONIX Lizenztypen: Neuer Code 20 für Lizenzen in ODRL Konvention

Alle Codewerte der Liste 218 - (License expression type)

| Codewert | Bezeichnung | Beschreibung | Version | Update |
|----------|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------|---------|--------|
| 01 | Human readable | Document (eg Word file, PDF or web page) Intended for the lay reader | 24 | 24 |
| 02 | Professional readable | Document (eg Word file, PDF or web page) Intended for the legal specialist reader | 24 | 24 |
| 10 | ONIX-PL | | 51 | 24 |
| 20 | ODRL | Open Digital Rights Language (ODRL) in JSON-LD format. Used for example to express TDM licenses using the W3C TDM Reservation Protocol | 65 | 65 |



KI Nutzung ausschließen - Lizenzdatei

```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://provider.com/policies/1",
  "assigner": {
    "uid": "https://provider.com",
    "vcard:fn": "Provider",
    "vcard:hasEmail": "mailto:contact@provider.com",
    "vcard:hasTelephone": "tel:+6175555555",
    "vcard:hasURL": "https://provider.com/tdm/licensing.html"
  },
  "permission": [{
    "action": "tdm:mime",
    "duty": [{
      "action": "obtainConsent"
    }]
  }]
}
```



KI Nutzung ausschließen - Assets

- Neu kann auch für die einzelnen Assets und Texte der Nutzung für TDM widersprochen werden:

```
<EpubUsageConstraint>
  <EpubUsageType>11</EpubUsageType>      <!-- TDM, in list 145 -->
  <EpubUsageStatus>03</EpubUsageStatus> <!-- is prohibited, list 146 -->
</EpubUsageConstraint>
```

- Auch hier kann jetzt eine “Lizenz” mitgegeben werden, die das Recht für TDM genauer regelt oder eine Lizenz dafür anbietet:

```
<EpubLicense>
  <EpubLicenseName>MyPublisher Asset TDM Policy</EpubLicenseName>
  <EpubLicenseExpression>
    <EpubLicenseExpressionType>20</EpubLicenseExpressionType> <!-- ODRL license, list 218 -->
    <EpubLicenseExpressionLink>https://www.mypublisher.com/.policies/assetpolicy.json
  </EpubLicenseExpression>
</EpubLicense>
```



KI Nutzung ausschließen - Assets

- Das Opt Out und die Lizenz werden direkt in SupportingResource und TextContent für jedes Asset einzeln angegeben

```
<SupportingResource>
  . . .
  <ResourceVersion>
    . . .
    <ResourceLink>https://www.mypublisher.com/readingguides/.../guide.pdf</ResourceLink>
    <EpubUsageConstraint>
      <EpubUsageType>11</EpubUsageType>      <!-- TDM -->
      <EpubUsageStatus>03</EpubUsageStatus>  <!-- is prohibited -->
    </EpubUsageConstraint>
    <EpubLicense>
      <EpubLicenseName>MyPublisher TDM License Policy</EpubLicenseName>
      <EpubLicenseExpression>
        <EpubLicenseExpressionType>20</EpubLicenseExpressionType>
        <EpubLicenseExpressionLink>https://www.mypublisher.com/.policies/mypolicy.json
        </EpubLicenseExpressionLink>
      </EpubLicenseExpression>
    </EpubLicense>
  </ResourceVersion>
  . . .
</SupportingResource>
```

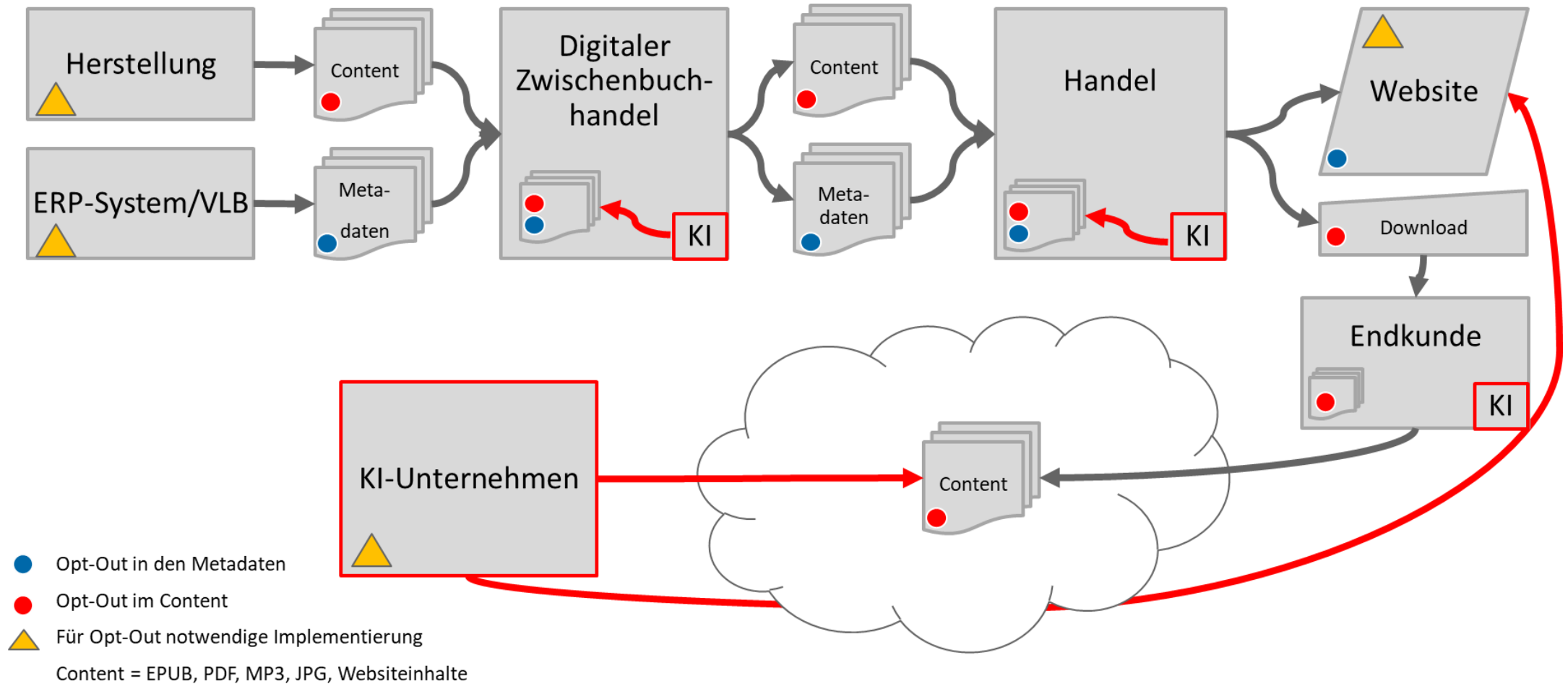


KI Nutzung ausschließen - ToDos

- Wichtig: Upgrade auf ONIX 3.1.1 erforderlich
- Service- und Software-Anbieter müssen die Angaben zu TDM und Lizenz in ihre ERP-, CMS- und MAM-Systeme aufnehmen
- Beispiel-Lizenzdateien bereitstellen
- TDM- und Lizenzangaben müssen in den ONIX Meldungen weitergegeben werden
- Verzeichnisse und Kataloge müssen die Kennzeichnungen annehmen, speichern und weitergeben
- Händler müssen auf Ihren Seiten die TDM-Restriktionen mit den Assets anzeigen und kennzeichnen
- Metadatenelemente für Assetdateien (epub, pdf, jpg, png, video und audio) müssen (weiter) entwickelt werden.



Technische Umsetzung des AI-Opt-out



KI generierten Content kennzeichnen

- ... und nicht von KI generierten Content ggf. ebenfalls kenntlich machen



Marsel van Oosten / Knesebeck - no_ai



KI generierten Content kennzeichnen

- Ausgangssituation
 - Fehlende Kennzeichnung von KI generiertem Content
 - Kein Standard für die Kennzeichnung
 - Qualitativ minderwertige KI-generierte E-Books bei Händlern
 - Desinformation
 - Vertrauensverlust
 - Verzerrung der Realität (durch unmarkierte Fake-Bilder)
- Ziel: Deklaration von KI-Content, um Handelspartnern und Endkund*innen den Ursprung des Produkts darzustellen
- Ausprägungswünsche: generated, under_collaboration, assisted, no_ai



no_ai



assisted

KI generierten Content kennzeichnen

- Wird KI generierter Content in Verlagspublikationen verwendet, sollte dieser entsprechend gekennzeichnet sein
- Die Kennzeichnung erfolgt im Contributor Composite
- Ein KI generiertes Cover könnte also so angezeigt werden:

```
<Contributor>  
  <SequenceNumber>2</SequenceNumber>  
  <ContributorRole>A36</ContributorRole>  <!-- cover artwork by -->  
  <UnnamedPersons>09</UnnamedPersons>  <!-- generative AI -->  
</Contributor>
```



KI generierten Content kennzeichnen

- Oder der Autor hatte KI Unterstützung:

```
<Contributor>
  <SequenceNumber>1</SequenceNumber>
  <ContributorRole>A01</ContributorRole>
  <PersonName>Elke Mustermann</PersonName>
</Contributor>
<Contributor>
  <SequenceNumber>2</SequenceNumber>
  <ContributorRole>Z01</ContributorRole>
  <UnnamedPersons>09</UnnamedPersons>
</Contributor>
```

<!-- author -->
<!-- human author -->

<!-- assisted by -->
<!-- generative AI -->



KI generierten Content kennzeichnen

- Verwendung für synthetische Stimmen in Hörbüchern mit Liste 19:

Alle Codewerte der Liste 19 - (Unnamed person(s))

| Codewert | Bezeichnung | Beschreibung | Version | Update |
|----------|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--------|
| 01 | Unknown | | 0 | 0 |
| 02 | Anonymous | Note that Anonymous can be interpreted as singular or plural. A real name can be provided using <AlternativeName> where it is generally known | 0 | 49 |
| 03 | et al | And others. Use when some but not all contributors are listed | 0 | 0 |
| 04 | Various | When there are multiple contributors, and none are listed individually. Use for example when the product is a pack of books by different authors | 2 | 49 |
| 05 | Synthesised voice – male | Use with Contributor role code E07 'read by', eg for audio books with digital narration | 8 | 49 |
| 06 | Synthesised voice – female | Use with Contributor role code E07 'read by', eg for audio books with digital narration | 8 | 49 |
| 07 | Synthesised voice – unspecified | Use with Contributor role code E07 'read by', eg for audio books with digital narration | 8 | 49 |
| 08 | Synthesised voice – based on real voice actor | Use with Contributor role code E07 'read by', eg for audio books with digital narration, and provide name of voice actor in <AlternativeName>. Only for use in ONIX 3.0 or later | 49 | 49 |
| 09 | AI (Artificial intelligence) | Use when the creator (of text, of images etc) is a generative AI model or technique. Note, can also be combined with the role 'assisted by'. Only for use in ONIX 3.0 or later | 62 | 62 |

Ausblick

Die Peergroup “**KI im ONIX**” wird alle aufgeführten Fälle UND jeden weiteren Bedarf in einem **Best Practice** festhalten!

Carsten Wehmeyer / VVA-Arvato
Daniel Gebühr / Apotheker Verlag
Isabell Jablonski / Droemer-Knaur
Julian Vogel / MVB
Sabrina Bressel / Kohlhammer Verlag
Sven Simoneit / Umbreit



Vielen Dank !

Weiterführende Informationen:

- Freigabe der ONIX Version 3.1.1:
<https://www.editeur.org/93/Release-3.0-and-3.1-Downloads/>
- Aspekte von KI Nutzung in ONIX Meldungen:
<https://www.editeur.org/files/ONIX%203/APPNOTE%20Aspects%20of%20AI%20in%20ONIX.pdf>
- Metadaten in epubs:
<https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/>
(This is not a W3C standard, but is a specification developed by a W3C Community Group and widely supported by organisations in the publishing industry. It may become a W3C recommendation in the future.)
- TDM Policy Dateien nach W3C Empfehlung:
<https://w3c.github.io/tdm-reservation-protocol/spec/#sec-policy>
- Aktuelle Beschäftigung des Börsenvereins mit dem Thema KI:
<https://www.boersenverein.de/digitaler-wissens-hub/alle-beitraege/detail-seite/grundsaeetze-zum-fairen-umgang-mit-ki-in-der-buchbranche/#c45048> (nur für Mitglieder)

