Memorandum

JENTIS GmbH

Verfasser/in	**Datum des Dokuments**
RA Tilman Herbrich (CIPP/E)	*19.07.2023, v1.1*

Projekt:

Datenschutzrechtliche Bewertung des Basiskonzepts „Synthetische User“

Executive Summary

Aus der Betrachtung des Basiskonzepts „Synthetische User“[1] folgt, dass Unternehmen die **JENTIS Data Capture Platform** (“DCP”) und den „JENTIS Essential Mode“[2] als Fallback-Lösung nutzen können, um das First-Party-Tracking so zu konfigurieren, dass die Anwendung der Ausnahmevorschriften vom Einwilligungserfordernis für den Endgerätezugriff (z. B. § 25 Abs. 2 TTDSG) im Einklang mit aufsichtsbehördlichen Positionen belastet werden können.

Durch Anwendung **mathematischer Verfahren** auf Basis von Nutzerdaten, für die eine Einwilligung vorliegt, und von reduzierten technisch notwendigen oder unbedingt erforderlichen mittels des JENTIS Essential Mode erfassten Nutzerdaten, **ermöglicht** JENTIS eine **statistische Analyse** bis 100% des **Nutzungsverhaltens** digitaler Angebote im reduzierten Umfang **ohne Abfrage einer Einwilligung.**

Beurteilungsgrundlage

1. Auswirkungen industrieller Entwicklungen und Consent-Rate auf Analyseverfahren

Die Messbarkeit des Nutzungsverhaltens digitaler Angebote beträgt im Durchschnitt 30% des Zielpublikums. Ausgehend von einer durchschnittlichen Consent-Rate und dem negativen Einfluss von **Ad-Blockern** und **Tracking-Blockadesystemen** von Browsern per Voreinstellung wie Safari (Intelligent Tracking Prevention „ITP“) und Firefox (Enhanced Tracking Prevention, „ETP“) auf die Messung des Nutzungsverhaltens mittels Third- und First-Party-Tracking wächst die Bedeutung an verlässlichen Messergebnissen. Seitdem Safari (v16+) auch in bestimmten Fällen serverseitige Cookies von Drittanbietern blockiert, die über eine First-Party-Domäne gesetzt werden, verschärft sich die fehlende Möglichkeit zur Analyse des Nutzungsverhaltens digitaler Angebote.

2. Erzeugung Synthetischer Nutzer mit JENTIS Synth-User-Engine & Server-Side-Tracking

Im Durchschnitt können bis zu 70% das Nutzerverhalten mangels Erteilung einer Einwilligung oder Blockierung der Nutzerbrowser nicht gemessen werden, so dass nur die Daten von 30% der Nutzer tatsächlich als Entscheidungsgrundlage für das Unternehmen verwendet werden. JENTIS ermöglicht die Erzeugung Synthetische User zur validen Pseudonymisierung von Nutzungsdaten bei aktiviertem JENTIS Essential Mode.[3] Hierzu werden anstatt der 70% echten, aber für die Analyse fehlenden Nutzer, ersatzweise Synthetische Nutzer,

[1] Basiskonzept “Erstellung von synthetischen Usern als Folge der kombinierten Verarbeitung von eingewilligten und unbedingt erforderlichen Daten im Sinne der ePrivacy und der DSGVO zum Zweck von Online Marketing.”, JENTIS GmbH, April 2023 (siehe Anhang).
[2] Informationen zum JENTIS Essential Mode siehe hier.
[3] *Ibid.*

d. h. künstliche Nutzer ohne Personen- oder Endgerätebezug generiert, um diese zusammen mit 30% der echten Nutzer zu messen und auf diese Weise 100% Sichtbarkeit im Analysetool sicherzustellen.

Die **JENTIS Synth-User Engine** wird anhand von historischen Daten sowie der Live-Tracking-Daten (sowohl von Consent- als auch Nicht-Consent-Nutzern) mittels der statistischen Methode aus der Mathematik „Imputation" die fehlenden Daten im Nicht-Consent-Nutzer Bereich neu erzeugen. In einem der Imputation vorangehenden Schritt werden die Nutzerdaten in beiden Gruppen (sowohl von Consent- als auch Nicht-Consent-Nutzern) anhand von Prädiktoren geclustert. Auf diese Weise können unvollständige Datensätze durch funktionale Zusammenhänge vervollständigt werden. Nach Anwendung des statistischen Verfahrens Imputation werden alle **Rohdaten** von **Nutzern ohne Einwilligung** mithilfe der JENTIS-Technologie vollständig **gelöscht**.

Für die **Erzeugung Synthetischer Nutzer** werden die künstlichen Datensätze mit den Datensätzen der realen Consent-Nutzer kombiniert, um neue Verhaltensdaten von nicht realen, künstlichen Nutzern zu generieren. Der Datensatz, bestehend aus realen und synthetischen Daten, kann über die **JENTIS Server-Side-Tracking-Technologie** als technischer Vorfilter an Analyse-, oder Marketing-Tools zu Analysezwecken übergeben werden.

Rechtliche Würdigung

Im Europäischen Datenschutzrecht hat sich das **Konzept** der **Synthetisierung von Daten**[4] als eine **Privacy Enhancing Technologie** (PET) fest etabliert.[5] Nach Ansicht des Europäischen Datenschutzbeauftragten (**EDPS**) und der Agentur der Europäischen Union für Cybersicherheit (**ENISA**) stellen synthetische Daten eine PET und in diesem Sinne eine **zusätzliche Schutzmaßnahme** für **Datenübertragungen** – auch in Drittländer ohne angemessenes Schutzniveau – dar.[6]

1. Definition Synthetischer Daten

Synthetische Daten werden auch „fake data" oder „künstliche Daten" (artificial data) genannt. Unabhängig von der Terminologie sind synthetische Daten auf grundlegender Ebene künstlich erzeugte Daten, die aus den Originaldaten erzeugt werden und die statistischen Eigenschaften dieser Originaldaten bewahren, ohne jedoch einen Bezug zu einer identifizierten oder identifizierbaren Person aufzuweisen.[7]

2. Datenschutzrechtliche Einordnung von synthetischen Daten

Aufgrund der Risiken einer Re-Identifizierung von betroffenen Personen[8] sind **synthetisierte Daten**, die aus realen Daten betroffener Personen generiert werden, **zumeist** nicht als anonyme Daten, sondern als **pseudonyme Daten** einzuordnen. Insbesondere dann, wenn synthetische Daten eine ausreichende strukturelle

[4] Vgl. zur Entwicklung und den Methoden EU-Commission, JRC Technical Report, 2022, S. 12 ff.
[5] EDPS, techsonar 2021-2022, S. 10; ENISA, Data Protection Engineering, 2022, S. 17.
[6] EDPS, techsonar 2021-2022, S. 10; Hintze/Emam, Can synthetic data help organizations respond to 'Schrems II'?, 2020.
[7] EDPS, techsonar 2021-2022, S. 10; López/Elbil, European Law Blog, On synthetic data: a brief introduction, 2022.
[8] EDPS, techsonar 2021-2022, S. 11.

Äquivalenz mit dem Originaldatensatz aufweisen oder wesentliche Eigenschaften oder Muster teilen, die eine Zuordnung auf reale Nutzer zulassen, ist von einer Pseudonymisierung i. S. d. Art. 4 Nr. 5 DSGVO auszugehen.[9]

Die **ENISA** verweist auf diverse Studien, wonach selbst bei einem vollständig anonymisierten Datensatz von Verkehrsdaten (mobility data) drei bis vier bekannte Datenpunkte ausreichen, um eine Re-Identifizierung durchzuführen.[10]

Die **Norwegische Datenschutzbehörde** hat im Juni 2021 ein Bußgeld gegen einen Verband aufgrund einer versehentlichen Veröffentlichung von Nutzerdaten mangels eines Testverfahrens einer Cloud-Lösung verhängt. In der Begründung verweist die Behörde darauf, dass die Nutzung von synthetischen Daten den Datenschutzvorfall hätte vermeiden können und daher die Nutzung stets zu empfehlen sei.[11]

4. Datenschutzrechtliche Beurteilung der Erzeugung Synthetischer User

(1) Die im JENTIS Basiskonzept „Synthetische User" beschriebene Erzeugung Synthetischer Nutzer[12] ist als robuste Maßnahme der Pseudonymisierung einzuordnen. Aufgrund der Ableitung von Verhaltensdaten aus realen Daten von Nutzern, die eine Einwilligung erteilt haben, und der Löschung von Rohdaten der Nutzer, die keine Einwilligung erteilt haben, ist der Vorgang der **Erzeugung Synthetischer User** lediglich als **Pseudonymisierung** i. S. d. Art. 4 Nr. 5 DSGVO einzustufen.

Die **Synthetisierung realer Rohdaten** wie die von Drittanbietern vergebene Client-ID oder User ID ist **unter denselben Bedingungen** wie die **Bildung** von **Hashwerten** aus realen Rohdaten als Pseudonymisierung i. S. v. Art. 4 Nr. 5 DSGVO einzuordnen.[13] Solange die an der Stelle von Client-IDs und User-IDs verwendeten **künstlichen Werte** für Drittanbieter **irreversibel** sind, die **Kollisionsresistenz** der aufbereiteten Datenparameter **sichergestellt** ist **und** die **IP-Adresse** des Nutzers durch künstliche Werte **ersetzt** wurde, ist unter Berücksichtigung der einhelligen Beurteilung zu Hashwerten mangels entgegenstehender Rechtsprechung von einer **DSGVO-konformen Pseudonymisierung** auszugehen.

(2) Für den Vorgang der Erzeugung Synthetischer User erfolgt **kein erneuter unmittelbarer Zugriff** auf **oder unmittelbare Speicherung** der Informationen in **Endgeräteressourcen** von Nutzern. Deshalb unterfällt der Vorgang als nachgelagerte Verarbeitungsphase im Einklang mit der **EuGH**-Entscheidung zum **One-Stop-Shop-**

[9] López/Elbil, European Law Blog, On synthetic data: a brief introduction, 2022; vgl. zu den Risiken der Re-Identifizierung allgemein Art. 29-Data Protection Working Party, WP 216, Opinion 5/2014 on Anonymisation Techniques, S. 11 sowie bei synthetischen Daten Stadler/Oprisanu/Troncoso, Synthetic Data – Anonymisation Groundhog Day, 2022, S. 4 ff.
[10] Vgl. ENISA, Data Protection Engineering, 2022, S. 10; vgl. etwa Gieselmann, Wie gängige Methoden zur Anonymisierung von Daten versagen, 2019.
[11] Norwegian Data Protection Authority (Datatilsynet), Press release of 15 June 2021.
[12] Basiskonzept "Erstellung von synthetischen Usern als Folge der kombinierten Verarbeitung von eingewilligten und unbedingt erforderlichen Daten im Sinne der ePrivacy und der DSGVO zum Zweck von Online Marketing.", JENTIS GmbH, April 2023 (siehe Anhang).
[13] Vgl. zum Hashing als valide Maßnahme zur Pseudonymisierung Schwartmann/Weiß, Draft for a Code of Conduct on the use of GDPR compliant pseudonymisation, 2019, v1.0, S. 26; ENISA, Pseudonymisation techniques and best practices, 2019, S. 33; ENISA, Data Pseudonymisation: Advanced Techniques & Use Cases, 2021, S. 12; Artikel 29-Data Protection Working Party, WP 216, Opinion 05/2014 in Anonymisation Techniques, S. 20.

Verfahren[14], der **Gesetzesbegründung** zum TTDSG[15] und der Ansicht des **EDPB**[16] ausschließlich den Regelungen der DSGVO.

(3) Soweit Nutzerdaten von Consent-Nutzern erfasst und zur Erzeugung Synthetischer User verwendet werden, liegt beiden Verarbeitungsvorgängen eine Einwilligung der Nutzer zugrunde. Die Daten der Consent-Nutzer werden möglichst genau gemessen und der Dat ensatz wird verwendet, um Prädiktoren zu bestimmen, die eindeutig sind, um die zustimmenden Nutzer zu extrapolieren. Die Prädiktoren legen die Datenparameter fest, die von den Nicht-Consent-Nutzer erhoben werden. Entsprechende Rohdaten werden vor Weitergabe der Daten an andere Systeme vollständig gelöscht.

Daten, die von **Nicht-Consent-Nutzern** erhoben werden, können auf der Grundlage der **Ausnahmeregelung** nach Art. 5 Abs. 3 S. 2 ePrivacy-RL[17] und § 25 Abs. 2 TTDSG[18] (bzw. TKG in Österreich) qualifiziert von der Einwilligung ausgenommen werden, sofern diese Daten (auch als "Prädiktoren" bezeichnet) nicht persistent sind und auf das beschränken, was technisch notwendig oder unbedingt erforderlich ist, um Nutzer mit Einwilligung hochzurechnen (durch Imputation).[19]

Als **Prädiktor** wird im Zusammenhang mit der JENTIS-Technologie für Synthetische Nutzer eine technische (in der Regel nicht personenbezogene) **Metrik** bezeichnet, die der Algorithmus benötigt, um die Nutzer zu bündeln ("clustern"), bevor er die Daten der nicht eingewilligten Nutzer mit den Daten der eingewilligten Nutzer imputiert, **um** den **Synthetischen Nutzer zu erzeugen**.

Beispiele für einen **Prädiktor**, der für das Clustering von Nutzern verwendet wird, können sein:

- welcher Browser verwendet wird;
- Dauer der Session;
- Zeitpunkt des Sessionbeginns;
- Anzahl der Pageviews;
- ob sich ein bestimmtes Produkt im Warenkorb befindet;
- Scroll-Rate auf der 1. Seite.

Die Prädiktoren werden von dem mathematischen Modell automatisch auf der Grundlage der von den Nutzern, die ihre ausdrückliche Einwilligung gegeben haben, verarbeiteten Daten extrahiert. Der qualitative Test wird

[14] EuGH, Urt. v. 15.06.2021 – C-645/19 – One-Stop-Shop, Rn. 74.
[15] BT-Drs. 19/27441, S. 38.
[16] EDPB, Stellungnahme 5/2019 zum Zusammenspiel zwischen der e-Datenschutz-Richtlinie und der DSGVO, S. 23.
[17] Art. 5 Abs. 3 S. 2 ePrivacy-RL: "Dies steht einer technischen Speicherung oder dem Zugang nicht entgegen, wenn der alleinige Zweck die Durchführung oder **Erleichterung** der Übertragung einer Nachricht über ein elektronisches Kommunikationsnetz ist oder, soweit dies unbedingt erforderlich ist, um einen vom Teilnehmer oder Nutzer ausdrücklich gewünschten Dienst der Informationsgesellschaft zur Verfügung zu stellen."
[18] § 25 Abs. 2 TTDSG: "(2) Die Einwilligung nach Absatz 1 ist nicht erforderlich,
1. wenn der alleinige Zweck der Speicherung von Informationen in der Endeinrichtung des Endnutzers oder der alleinige Zweck des Zugriffs auf bereits in der Endeinrichtung des Endnutzers gespeicherte Informationen die Durchführung der Übertragung einer Nachricht über ein öffentliches Telekommunikationsnetz ist oder
2. wenn die Speicherung von Informationen in der Endeinrichtung des Endnutzers oder der Zugriff auf bereits in der Endeinrichtung des Endnutzers gespeicherte Informationen unbedingt erforderlich ist, damit der Anbieter eines Telemediendienstes einen vom Nutzer ausdrücklich gewünschten Telemediendienst zur Verfügung stellen kann."
[19] *Siehe*: Imputation.

angewandt, um die Prädiktoren zu bestimmen, die eine genaue Clusterbildung ermöglichen würden. Es gibt eine technische Möglichkeit, bestimmte Datenpunkte von der Einstufung als Prädiktor auszuschließen.

Die kurzweilige **Verarbeitung** der **Prädiktoren** für die Erstellung von Gruppen, bestehend Synthetischen Nutzern zur Sicherstellung einer Reichweitenmessung für die fehlerfreie Auslieferung digitaler Angebote wie Websites kann im Einklang mit aufsichtsbehördlichen Positionierungen **auf** die **Ausnahmeregelung** in § 25 Abs. 2 Nr. 2 TTDSG **gestützt** werden. Im Rahmen der Erforderlichkeitsprüfung für die Belastbarkeit von § 25 Abs. 2 Nr. 2 TTDSG nach Maßgabe des EDPS „Necessity Toolkit“[20] wurden die Kriterien der Datenschutzkonferenz (DSK) aus der Orientierungshilfe für Anbieter:innen von Telemedien[21] vollständig berücksichtigt.

Maßgebliche Kriterien für die Bestimmung der **unbedingten Erforderlichkeit** sind Ansicht der DSK:

- **Zeitpunkt der Speicherung**: Das Auslesen der Prädiktoren-Daten beginnt ab dem Zeitpunkt, zu dem der Nutzer auf die Website aufruft.
- **Dauer der Speicherung**: Die Dauer der Speicherung der Prädiktoren-Daten ist nicht persistent und erfolgt lediglich für den Zeitraum der Imputation und Bildung der Cluster für die Nicht-Consent-Nutzern, wobei der genaue Zeitraum vom Nutzeraufkommen der Website abhängt. Der Zeitraum kann wenige Minuten bis mehrere Stunden in Anspruch nehmen. Die Prädiktoren-Daten werden vom Endgerät ausgelesen und bis zu dem Zeitpunkt gespeichert, zu dem synthetische Nutzer auf der Grundlage der durch Prädiktoren ermittelten Cluster erstellt werden. Im Anschluss des Vorgangs werden sämtliche Rohdaten von Nicht-Consent-Nutzern gelöscht.
- **Inhalt der Daten**: Die Prädiktor-Daten sind eine sehr begrenzte Anzahl von technischen Metriken, die von einem Algorithmus sorgfältig gefiltert werden. Die Prädiktoren werden automatisch auf der Grundlage der von den Nutzern verarbeiteten Daten extrahiert, die ihre ausdrückliche Zustimmung gegeben haben. Die Prädiktoren werden verwendet, um eine möglichst genaue Clusterbildung zu ermöglichen. Es besteht die technische Möglichkeit, bestimmte Datenpunkte von der Einstufung als Prädiktor auszuschließen. Präditkoren können z.B. folgende nicht personenbezogene Endgeräteinformationen sein: session duration, time point of starting session, amount of page views, if a certain product is in the cart und scroll rate on the 1rst page)
- Auslesbarkeit der Informationen: Die Daten werden nur für einen sehr begrenzten Zeitraum vom Anbieter des Telemediendienstes gelesen (1st Party) und dann endgültig gelöscht. Es werden keine Drittanbieter für die Bildung Synthetischer Nutzer eingesetzt. Drittanbieter haben zu keinem Zeitpunkt Zugriff auf Endgeräte der Nutzer.

Die Reichweitenmessung für die bedarfsgerechte und fehlerfreie Darstellung digitaler Angebote wie Websites oder Apps erfordert eine Nutzungsanalyse auf Grundlage des gesamten Zielpublikums und wird von der DSK explizit als Beispiel für eine zulässige Reichweitenmessung genannt (Orientierungshilfe für Anbieter:innen von Telemedien (v1.1. S. 28 f.).

[20] Bewertung der Notwendigkeit von Maßnahmen, die das Grundrecht auf den Schutz personenbezogener Daten einschränken: Ein Toolkit, 11. April 2017
[21] DSK, Orientierungshilfe für Anbieter:innen von Telemedien, v1.1, S. 27 - S. 30.

(4) Aufgrund der vorgenommenen Pseudonymisierung lässt sich für die Übermittlung der synthetisierten Daten an Drittanbieter beim Server-Side-Tracking die Rechtsgrundlage gemäß Art. 6 Abs. 1 S. 1 lit. f) DSGVO (berechtigtes Interesse) in vertretbarer Weise anwenden.

(5) Für die **rechtskonforme Nutzung** Synthetischer User sind **folgende Maßnahmen** zu **empfehlen**:

- Konfiguration der JENTIS DCP bei aktiviertem JENTIS Essential Mode durch Kunden;
- Dokumentation der Interessenabwägung nach Art. 6 Abs. 1 S. 1 lit. f) DSGVO[22] durch ein sog. LIA (Legitimate Interests Assessment).
- Anpassung Einwilligungstexte in CMP in Bezug auf die Nutzung zur Erstellung Synthetischer User;
- Transparente Informationen in den Datenschutzinformationen auf der Website;
- Abschluss und Dokumentation Auftragsverarbeitungsvertrag mit JENTIS.

[22] EDPB, Leitlinien für Transparenz gemäß der Verordnung 2016/679, WP 260, rev.01, Anhang.