Das private Palantir

KI-Agenten, öffentliche Daten und das Ende praktischer Obskurität

Autor: Benjamin Metzig

Projekt: Wissenschaftswelle.de

Datum: 12. Juni 2026

Publikationsform: Deutsche PDF-Fassung

Forschungsbericht für öffentliche Debatte und wissenschaftliche Weiterentwicklung.

Das private Palantir

Forschungsbericht über KI-Agenten, öffentliche Daten und das Ende praktischer Obskurität

Stand: 12. Juni 2026
Zweck: Bereitstellung einer belastbaren Forschungs- und Diskussionsgrundlage für öffentliche, journalistische und politische Debatten über das Risiko KI-gestützter Personenprofilierung durch private, nichtstaatliche Akteure.
Kernbefund: Umfassende private Überwachung sollte nicht als bereits gegenwärtige Alltagsrealität behauptet werden. Belastbar ist aber: Die technischen, organisatorischen und ökonomischen Bedingungen für gezielte automatisierte Personenprofilierung aus öffentlichen Daten verbessern sich schnell. Dadurch sinkt die Schwelle für Stalking, Doxxing, Einschüchterung, Betrug, informelles Arbeitgeber-Screening und lokale Formen sozialer Macht.

Autor: Benjamin Metzig
Projektkontext: Wissenschaftswelle.de, unabhängiges Wissens- und Wissenschaftsmagazin, Deutschland
Projektseite: https://www.wissenschaftswelle.de/
Empfohlene Zitierweise: Metzig, B. (2026). *Das private Palantir: Forschungsbericht über KI-Agenten, öffentliche Daten und das Ende praktischer Obskurität*. Wissenschaftswelle.de.
Transparenzhinweis: Dieser Bericht wurde von Benjamin Metzig für Wissenschaftswelle.de entwickelt. Redaktionelle Verantwortung, Quellenbewertung, Schlussfolgerungen und Veröffentlichungsentscheidung liegen bei Benjamin Metzig.

Abstract

Öffentliche personenbezogene Informationen sind nicht gleichbedeutend mit frei und grenzenlos nutzbaren Informationen. Viele Datenschutz- und Privatsphärenschutzwirkungen beruhen auf praktischer Obskurität: Zerstreuung, Aufwand, Kontext, Mehrdeutigkeit, Zeitablauf und begrenzte Verknüpfbarkeit. Dieser Bericht argumentiert, dass KI-Agenten, multimodale Modelle, lange Kontextfenster, Speicherarchitekturen, Werkzeugnutzung und lokal verfügbare offene Modelle diese Schutzschicht schwächen, weil sie die Kosten des Sammelns, Verknüpfens, Speicherns und Interpretierens öffentlicher Spuren über identifizierbare Personen senken. Der Bericht schlägt **Private Intelligence Automation** als analytische Risikokategorie vor: die nichtstaatliche, automatisierte, wiederholbare und oft intransparente Aggregation öffentlicher oder halböffentlicher personenbezogener Informationen für private Zwecke. Die Evidenz belegt keine bereits verbreitete totale Privatüberwachung. Sie belegt aber relevante technische Fähigkeiten, angrenzende Schadensfelder und Governance-Lücken. Doxxing, technologiegestützte Gewalt, Social Engineering, informelles Beschäftigtenscreening, Datenbrokerage und Chilling Effects liefern empirische Anker für die Risikologik. Eine balancierte Pilotstudie mit 125 öffentlichen kommunalen PDF-Publikationen aus fünf deutschen Städten zeigt zusätzlich, dass wiederkehrende Amts- und Stadtblätter dauerhafte

Kombinationen aus Kontaktmustern, Rollen, Daten, schul-/jugend-/familienbezogenen Begriffen und Metadaten enthalten können. Der Bericht schlägt eine nicht-operative Forschungs- und Politikagenda vor: Prävalenzmessung, synthetische Agenten-Evaluationen, institutionelle Expositionsaudits, vergleichende Governance und Schutzmaßnahmen gegen wiederholte personenbezogene Aggregation.

Schlagwörter: KI-Agenten; öffentliche Daten; praktische Obskurität; Datenschutz; Doxxing; digitale Gewalt; Datenbroker; OSINT; Open-Weight-Modelle; Personenprofilierung; KI-Governance.

Executive Summary

Dieser Bericht untersucht die These, dass KI-Agenten, multimodale Modelle, Web- und Computer-Use-Systeme, lange Kontextfenster, Speicherarchitekturen und offene bzw. lokal ausführbare Modelle eine neue Risikokategorie entstehen lassen: **Private Intelligence Automation**. Gemeint ist die nichtstaatliche automatisierte Sammlung, Verknüpfung, Speicherung und Bewertung öffentlich zugänglicher personenbezogener Informationen.

Die Gefahr liegt nicht in einer einzelnen magischen KI-Fähigkeit. Sie entsteht aus der Kopplung bereits sichtbarer Bausteine: Rechercheagenten, Dokument- und Bildanalyse, Werkzeugnutzung, Browserbedienung, Netzwerkinferenz, Profilbildung, Speicher, Wiederholung und fallende Kosten. Verstreute öffentliche Daten können zu privaten Dossiers werden; private Dossiers können zu sozialem Druck, Einschüchterung oder unfairen Entscheidungen werden.

Die Evidenz für technische Bausteine ist stark. Anbieter und Forschung dokumentieren Systeme mit mehrstufiger Recherche, Werkzeugnutzung, Computerbedienung, multimodaler Verarbeitung, langen Kontextfenstern und offenen Modellgewichten. Gleichzeitig zeigen Benchmarks wie OSWorld, WebArena und Online-Mind2Web, dass Agenten weiterhin fehleranfällig, langsam und in realen Web- und Computerumgebungen begrenzt zuverlässig sind. Diese Grenzen dämpfen übertriebene Behauptungen, beseitigen das Risiko aber nicht: Für viele Missbrauchsszenarien reicht Teilautomatisierung.

Die Evidenz für angrenzende soziale Schäden ist moderat bis stark. Europol, BKA und FBI dokumentieren bereits, dass KI Betrug, Social Engineering und Cybercrime professionalisieren kann. HateAid und die Technische Universität München zeigen, dass digitale Gewalt politisches Engagement in Deutschland beeinträchtigt. Datenschutzrechtlich sind öffentlich verfügbare personenbezogene Daten nicht frei für beliebige Weiterverarbeitung; DSGVO, EDPB-Leitlinien und Orientierungshilfen der Datenschutzkonferenz betonen Rechtsgrundlage, Zweckbindung, Transparenz, Datenminimierung und Schutzbedürftigkeit personenbezogener Verarbeitung.

Die wichtigste Unsicherheit bleibt die Prävalenz. Es gibt noch wenig robuste empirische Daten darüber, wie häufig private Akteure KI bereits systematisch für Personenrecherche, Stalking, Doxxing oder informelle Profilbildung nutzen. Die angemessene Bewertung ist deshalb nüchtern: Nicht die Behauptung, umfassende private Überwachung sei schon alltäglich, sondern die Feststellung, dass die Schwelle für gezielte Profilierung sinkt, dass die Schadensformen plausibel sind und dass die Governance-Lücken groß genug sind, um jetzt Forschung und Schutzmaßnahmen zu entwickeln.

Beitrag und Geltungsbereich

Der Bericht leistet fünf Beiträge:

1. **Begrifflicher Beitrag:** Er definiert Private Intelligence Automation als analytische Risikokategorie: nichtstaatliche, automatisierte, wiederholbare und oft intransparente personenbezogene Aggregation aus öffentlichen oder halböffentlichen Spuren.
2. **Theoretischer Beitrag:** Er verbindet praktische Obskurität, kontextuelle Integrität und Online-Obskurität mit spezifischen Fähigkeiten agentischer KI: Suche, Synthese, multimodale Analyse, Speicher, Werkzeugnutzung und Aktualisierung.
3. **Evidenzbezogener Beitrag:** Er trennt Evidenz für technische Fähigkeiten, Evidenz für soziale Schäden und Evidenz für tatsächliche Verbreitung. Das verhindert Alarmismus ebenso wie vorschnelle Entwarnung.
4. **Governance-Beitrag:** Er schlägt eine nicht-operative Forschungs- und Politikagenda vor, die Schäden untersuchen und reduzieren kann, ohne missbrauchsfähige Workflows zu veröffentlichen oder legitime OSINT-, Journalismus-, Cybersicherheits- und Menschenrechtsarbeit zu beschädigen.
5. **Empirischer Pilotbeitrag:** Er integriert eine balancierte Auditstudie mit 125 öffentlichen kommunalen PDF-Publikationen aus fünf Städten, um zu prüfen, ob institutionelle Publikationspraktiken aggregierbare Spuren erzeugen, ohne einzelne Personen zu untersuchen.

Der Bericht ist enger als allgemeine KI-Sicherheit und breiter als eine einzelne Schadenskategorie. Er behauptet nicht, dass KI Stalking, Doxxing, Betrug oder Diskriminierung neu erfunden habe. Er argumentiert, dass agentische KI Kosten, Geschwindigkeit, Wiederholbarkeit und scheinbare Autorität bereits bestehender Praktiken verändert.

Zentrale Befunde

1. **Die technischen Bausteine existieren.** Mehrstufige Webrecherche, Dokumentenanalyse, Bildverstehen, Tool-Calling, Computer-Use, lange Kontexte und Agenten-Harnesses sind dokumentierte Fähigkeiten, auch wenn ihre Qualität stark variiert.
2. **Die Schadenslogik erfordert keine perfekte Überwachung.** Falsche, selektive oder unsichere Profile können in Arbeitskontexten, lokalen Konflikten, politischer Einschüchterung, Stalking und Betrug Schaden verursachen, bevor sie verifiziert werden.
3. **Praktische Obskurität ist eine reale Schutzschicht.** Informationen können öffentlich sein und dennoch durch Zerstreuung, Aufwand, Kontext und Vergessen funktional geschützt bleiben. Agentische Systeme reduzieren genau diese Reibung.
4. **Open-Weight-Modelle sind ein eigenständiger Governance-Faktor.** Sie sind für Forschung, Wettbewerb und digitale Souveränität wichtig, erschweren aber zentrale Missbrauchskontrolle, weil lokale und angepasste Nutzung schwerer beobachtbar ist.

5. **Das Risiko betrifft nicht nur öffentliche Personen.** Relevante Kontexte sind Trennung und Stalking, politisch engagierte Personen, Journalisten, lokale Amtsträger, Beschäftigte, Bewerber, Minderjährige, queere Menschen, kleine Gemeinschaften und Personen mit sichtbaren beruflichen oder sozialen Rollen.
6. **Regulierung hat blinde Flecken.** DSGVO und EU AI Act liefern wichtige Prinzipien, erfassen private, informelle, lokale und schwer nachweisbare Profilierung aber nur teilweise.
7. **Die Forschungslücke ist selbst ein Risiko.** Ohne empirische Daten zu Nutzung, Schäden, Täterprofilen und wirksamen Gegenmaßnahmen bleibt die Debatte anfällig für Panik und für vorschnelle Entwarnung.

Evidenzklassen

Klasse	Bedeutung	Verwendung im Bericht
Etabliert	Gestützt durch Primärquellen, amtliche Dokumentation, Rechtsquellen, wissenschaftliche Arbeiten, Benchmarks oder robuste Lageberichte.	Technische Fähigkeiten, Datenschutzprinzipien, bekannte digitale Gewalt, Cybercrime-Trends.
Plausibel	Gestützt durch etablierte Bausteine und nachvollziehbare Kausalketten, aber ohne direkte Prävalenzdaten zum exakten Szenario.	Verbindung zwischen Agentenfähigkeiten und privater Personenprofilierung.
Unsicher	Relevante Indikatoren existieren, aber Datenbasis, Häufigkeit, Effekt oder Zeithorizont sind unklar.	Verbreitung privater agentischer Dossiers, Wirksamkeit von Schutzmaßnahmen.
Spekulativ	Möglich, aber derzeit nicht robust genug für starke Behauptungen.	Normalisierung umfassender lokaler Profilierungsagenten in drei bis fünf Jahren.

Hypothese und Forschungslogik

H1: In den nächsten ein bis drei Jahren steigt das Risiko nichtstaatlicher KI-gestützter Personenprofilierung aus öffentlichen Daten deutlich, weil technische Fähigkeiten, Automatisierung und fallende Kosten gemeinsam die Schutzwirkung praktischer Obskurität schwächen.

Teilhypothese	Evidenzklasse	Kurzbegründung
Agenten können Recherche, Extraktion, Werkzeugnutzung und Synthese bereits teilweise automatisieren.	Etabliert	Deep Research, ChatGPT Agent, Computer Use, Tool-Use-Dokumentation und Web-Agent-Benchmarks.
Agenten bleiben unzuverlässig, aber Teilautomatisierung genügt für viele Missbrauchsszenarien.	Etabliert/plausibel	Stalking, Betrug und reputative Schäden benötigen keine perfekte Vollautomatisierung.
Open-Weight-Modelle senken Kontroll- und Kostenbarrieren.	Etabliert/plausibel	AI Index, OECD, Llama, DeepSeek, Qwen.
Öffentliche Daten verlieren durch automatisierte Verknüpfung einen Teil ihrer sozialen Schutzwirkung.	Etabliert/plausibel	Praktische Obskurität, kontextuelle Integrität, Datenschutzrecht, Datenbrokerage.

Teilhypothese	Evidenzklasse	Kurzbegründung
Die Verbreitung privater KI-Dossiers ist noch unzureichend gemessen.	Unsicher	Fallzahlen, Tätertypologien und Langzeitstudien fehlen.

Kausalkette

Stufe	Beschreibung
1. Öffentliche und halböffentliche Spuren	Öffentliche Register, institutionelle Seiten, alte PDFs, Fotos, Veranstaltungslisten, berufliche Profile, Social-Media-Spuren.
2. KI-Recherche und Dokumentenanalyse	Suche, Extraktion, OCR, Zusammenfassung und Quellenvergleich reduzieren Sammelaufwand.
3. Normalisierung	Namen, Orte, Rollen, Daten, Organisationen und wiederkehrende Identifikatoren werden vergleichbar.
4. Inferenz	Hypothesen über Person, Umfeld, Interessen, Beziehungen, Verwundbarkeiten oder Routinen entstehen.
5. Speicherung und Aktualisierung	Episodische Recherche wird zu einer Datei, die erneut aufgerufen oder aktualisiert werden kann.
6. Nutzung	Profile, Dossiers oder Netzwerkkarten können Stalking, Doxxing, Betrug, Diskriminierung oder Einschüchterung unterstützen.
Querschnittstreiber	Fallende Kosten und lokale/offene Systeme erhöhen Machbarkeit und reduzieren zentrale Aufsicht.

Die Logik ist bewusst nicht-operativ. Sie beschreibt keine Anleitung, sondern eine abstrakte Risikokette.

Methodischer Rahmen

Der Bericht ist eine synthetische Risikoanalyse, keine Prävalenzstudie. Er nutzt eine strukturierte narrative Synthese: technische Evidenz, Rechtsquellen, Policy-Berichte, Datenschutztheorie und dokumentierte Schadensfelder werden zusammengeführt, um zu prüfen, ob eine neue Risikokategorie entsteht.

Forschungsfragen

1. Welche KI-Fähigkeiten sind für automatisierte Aggregation öffentlicher personenbezogener Informationen relevant?
2. Wie zuverlässig sind diese Fähigkeiten heute, und welche Grenzen sind für die Risikoabschätzung wichtig?
3. Welche bestehenden Schadensfelder könnten durch KI-gestützte Aggregation verstärkt werden?
4. Welche rechtlichen, institutionellen und produktbezogenen Governance-Rahmen adressieren das Risiko bereits, und wo liegen praktische Lücken?
5. Welche empirischen Studien wären nötig, um Prävalenz, betroffene Gruppen und wirksame Interventionen zu messen?

Quellengrundlage

Quellentyp	Beispiele	Rolle	Hauptgrenze
Anbieter- und Modelldokumentation	OpenAI, Anthropic, Google, Microsoft, Meta, Mistral, DeepSeek, Qwen	Belegt aktuelle Fähigkeiten.	Anbieter betonen oft Fähigkeit stärker als Missbrauch.
Benchmarks und wissenschaftliche Evaluationen	OSWorld, WebArena, Online-Mind2Web, AI Agents That Matter	Belegen Grenzen, Fehleranfälligkeit und Transferprobleme.	Benchmarks entsprechen nicht immer missbräuchlichen Low-Stakes-Workflows.
Recht und Regulierung	DSGVO, EDPB, DSK, EU AI Act, NIST AI RMF	Definieren Prinzipien personenbezogener Verarbeitung und Risikosteuerung.	Rechtslage ist jurisdiktionsabhängig.
Lageberichte und Zivilgesellschaft	Europol, BKA, FBI IC3, HateAid/TUM, Privacy International	Verknüpfen Mechanismus mit realen Schadensfeldern.	Messen meist nicht spezifisch agentische Personenaggregation.
Datenschutz- und Informationstheorie	Praktische Obskurität, kontextuelle Integrität, OSINT-Ethik	Begründet, warum Öffentlichkeit nicht freie Nutzung bedeutet.	Erfordert empirische Operationalisierung.

Terminologie

Praktische Obskurität bezeichnet die Bedingung, dass Informationen öffentlich zugänglich, aber durch Zerstreuung, Aufwand, Kontext, Alter, Mehrdeutigkeit oder fehlende Verknüpfbarkeit funktional geschützt sind.

Kontextuelle Integrität beschreibt Privatsphäre nicht als bloße Geheimhaltung, sondern als Einhaltung kontextabhängiger Informationsnormen: Wer darf welche Information in welchem Kontext und zu welchem Zweck nutzen?

KI-Agent meint hier ein System, das Ziele in Teilaufgaben zerlegen, Werkzeuge nutzen, Zwischenergebnisse verarbeiten, gegebenenfalls speichern und iterativ vorgehen kann.

Harness bezeichnet die Arbeitsumgebung um ein Modell: Speicher, Werkzeuge, Wiederholungen, Evaluation, Regeln, Logging und menschliche Freigaben. Für das private Palantir ist der Harness oft wichtiger als das Einzelmodell.

Private Intelligence Automation meint die automatisierte Sammlung, Verknüpfung, Speicherung und Bewertung öffentlicher personenbezogener Informationen durch nichtstaatliche Akteure. Der Begriff ist analytisch, nicht bereits rechtlich etabliert.

Agentic Stalking meint KI-gestütztes Stalking, das persistent, speicherbasiert, aktualisierend und vergleichend arbeitet. Auch dies ist ein analytischer Vorschlag, kein bestehender Straftatbestand.

Das private Palantir als Mechanismus

Das private Palantir ist keine einzelne Technologie. Es entsteht, wenn öffentliche Spuren, KI-Fähigkeiten, Speicherung, wiederholte Aktualisierung und ein schädlicher oder unverhältnismäßiger Zweck kombiniert werden.

Die zentrale Besonderheit ist die Kombination: **öffentliche Spuren + automatisierte Verknüpfung + Speicher oder Wiederholung + privater Zweck + geringe Anfechtbarkeit**.

Diese Kombination unterscheidet sich von klassischem OSINT, Datenbrokerage, Social-Media-Screening und staatlicher Überwachung. Sie kann Elemente all dieser Felder enthalten, ist aber oft informeller, privater, schwerer nachweisbar und weniger anfechtbar.

Technische Fähigkeitslage

Die technische Seite entwickelt sich entlang mehrerer Linien:

- **Websuche und Deep Research:** Systeme können mehrstufig recherchieren, Quellen zusammenfassen und PDF- oder Webinhalte verarbeiten.
- **Dokumenten- und Tabellenanalyse:** Modelle extrahieren Namen, Rollen, Daten, Kontaktmuster und Organisationen aus heterogenen Dokumenten.
- **Multimodalität:** Bilder, Screenshots und Layouts können beschrieben und in Kontext gesetzt werden.
- **Tool-Use und Computer-Use:** Agenten können Browser, Dateien, Formulare und externe Werkzeuge bedienen, bleiben aber fehleranfällig.
- **Lange Kontexte und Speicher:** Größere Kontextfenster und Retrieval-Systeme erleichtern fortlaufende Profile.
- **Open Weight und lokale Ausführung:** Zentralisierte Anbieterregeln werden weniger wirksam, wenn Workflows lokal oder in privaten Harnesses laufen.
- **Kostenentwicklung:** Sinkende Inferenzkosten und stärkere kleinere Modelle erhöhen die Skalierbarkeit.

Die relevante Frage lautet daher nicht: “Kann ein Agent zuverlässig alles über eine Person wissen?” Sondern: “Kann KI den Aufwand senken, genug Informationen zu sammeln, zu speichern und zu narrativieren, um Entscheidungen, Sicherheit oder Verhalten einer Person zu beeinflussen?” Diese Frage ist zunehmend mit Ja zu beantworten.

Schadensfelder

Stalking und coercive control: In Trennungskontexten können öffentliche Spuren genutzt werden, um Nähe, Kontrolle oder Einschüchterung zu erzeugen. KI senkt Aufwand und erhöht Wiederholbarkeit.

Doxxing und politische Einschüchterung: Journalisten, Aktivisten, lokale Amtsträger, Wissenschaftskommunikatoren und Engagierte sind sichtbar. Digitale Gewalt kann politische Beteiligung

bereits heute abschrecken; KI-gestützte Aggregation kann Zielgenauigkeit erhöhen.

Betrug und Social Engineering: Öffentliche Details erhöhen Glaubwürdigkeit personalisierter Betrugsversuche. KI kann Sprache, Zeitpunkt und Vorwand an Kontextspuren anpassen.

Arbeitswelt und informelle Diskriminierung: Arbeitgeber, Recruiter oder Dienstleister könnten öffentliche Spuren nutzen, um politische Ansichten, Gesundheit, Familienstatus, Aktivismus oder soziale Herkunft zu inferieren.

Lokale reputative Schäden: In kleinen Gemeinschaften können alte Fotos, Rollen, Vereinskontakte, Spenden, Veranstaltungen oder missverstandene Aussagen zu einer Dossier-Erzählung zusammengesetzt werden.

Chilling Effects: Die bloße Erwartung, dauerhaft aggregierbar zu sein, kann öffentliche Rede, politisches Engagement, Journalismus, Wissenschaftskommunikation und Vereinsleben verändern.

Pilotstudie: Lokale institutionelle Exposition

Zur Ergänzung der konzeptuellen Synthese enthält dieser Bericht eine balancierte empirische Pilotstudie. Sie fragt, ob wiederkehrende kommunale Publikationen aggregierbare Spuren erzeugen, ohne dass eine sensible Datenbank verletzt oder eine einzelne Person untersucht wird.

Forschungsfrage der Pilotstudie

Wie aggregierbar sind wiederkehrende öffentliche kommunale PDFs über vergleichbare städtische Publikationskontexte hinweg, und welche dokumentbezogenen Merkmale erhöhen das Risiko, dass öffentliche Spuren später in private Intelligenz überführt werden können?

Korpus und Ethik

Die Pilotstudie nutzt 125 öffentlich zugängliche kommunale PDF-Publikationen aus Heidelberg, Mannheim, Leipzig, Frankfurt am Main und Nürnberg. Der Korpus ist zwischen Städten balanciert, aber nicht repräsentativ für alle lokalen Institutionen. Jede Stadt trägt 25 live-validierte offizielle serielle PDFs bei, vor allem Amtsblätter, Stadtblätter und amtliche Bekanntmachungsformate.

Die Studie ist ein institutionelles Expositionsaudit, keine Personenrecherche. Es wurden keine Personen gezielt untersucht, keine Personenprofile erstellt, keine Namen, Rohtexte oder Roh-PDFs als Publikationsmaterial gespeichert und keine Bild- oder Gesichtsanalyse durchgeführt. Schul-, Jugend-, Kinderbetreuungs-, Bildungs- und Familienbegriffe wurden ausschließlich als dokumentbezogene Kontextindikatoren gezählt.

Indikatoren

Die Auswertung erfasste nur dokumentbezogene Aggregatindikatoren:

- Dichte namensähnlicher Kandidaten pro 10.000 Zeichen;
- E-Mail- und Telefonnummernmuster;
- Datumsdichte;

- Rollentermdichte;
- Dichte schul-/jugend-/familienbezogener Begriffe;
- Archiv-/Jahreskontext;
- PDF-Autor-/Creator-Metadaten;
- Dokumenttyp;
- Extraktionsqualität.

Der Aggregierbarkeitsscore ist ein heuristischer Dokumentindex. Er ist keine rechtliche Bewertung und kein Score für Personen. Er schätzt, wie viel Reibung ein Dokument für spätere Aggregation entfernt.

Ergebnisse

Alle 125 Dokumente wurden erfolgreich verarbeitet. Kein Dokument wurde manuell als dediziert minderjährigenbezogener Kontext markiert, aber schul-, jugend-, kinderbetreuungs-, bildungs- und familienbezogene Begriffe trugen in vielen Dokumenten zur Kontextkomponente bei. Alle 125 Dokumente enthielten E-Mail- und Telefonnummernmuster; 79 enthielten PDF-Autorenmetadaten. Der mittlere Aggregierbarkeitsscore lag bei 8,92, der Median bei 9, das Maximum bei 13. Die Extraktionsqualität war bei allen 125 Dokumenten hoch.

Risikostufe	Dokumente
Hoch	48
Sehr hoch	77

Stadt	Dokumente	Mittelwert	Median	Hoch	Sehr hoch	Autor-Metadaten
Frankfurt am Main	25	10,16	11	6	19	25
Heidelberg	25	7,68	8	23	2	0
Leipzig	25	9,28	9	0	25	25
Mannheim	25	8,80	9	8	17	25
Nürnberg	25	8,68	9	11	14	4

Die Pilotstudie zeigt nicht, dass einzelne Dokumente rechtswidrig oder schädlich sind. Sie zeigt, dass legitime öffentliche Publikation und spätere maschinelle Aggregierbarkeit auseinanderfallen können. Ein Dokument kann zum Zeitpunkt der Veröffentlichung angemessen sein und später dennoch Baustein einer Profilierung werden, wenn Suche, OCR, lange Kontexte, Speicher und wiederholte Aktualisierung kombiniert werden.

Grenzen der Pilotstudie

Die Stichprobe ist zwischen fünf Städten balanciert, aber auf serielle kommunale Publikationen begrenzt. Sie erfasst nicht die gesamte Vielfalt von Schulen, Vereinen, Veranstaltungsprogrammen, Sport, Kultur und Formularen. Die Indikatoren sind heuristisch und enthalten Fehlpositive. Namensähnliche Kandidaten sind keine verifizierten Namen. Bilder, Scans, Bildunterschriften,

Suchmaschinenindexierung, robots.txt, Plattform-Caching und Webarchive wurden nicht untersucht. Es wird kein kausaler Zusammenhang zu realem Missbrauch behauptet.

Für eine reine PDF-Veröffentlichung wird die Pilotstudie durch einen separaten PDF-Datenanhang begleitet. Dieser enthält die öffentliche Quellenliste, die Korpuszusammenfassung, stadtbezogene Ergebnisse und dokumentbezogene Aggregatindikatoren. Roh-PDFs, Rohtexte, Namen und Personenprofile sind nicht Teil des Publikationspakets.

Open-Weight-Modelle als Governance-Faktor

Offene und lokal ausführbare Modelle sind nicht “das Problem”. Sie haben legitime Gründe: Forschung, Transparenz, Wettbewerb, digitale Souveränität, lokale Verarbeitung, Datenschutz durch On-Device-Nutzung und geringere Abhängigkeit von wenigen Anbietern. Der Risikofaktor entsteht aus denselben Eigenschaften:

1. Missbrauch muss nicht durch zentrale Anbieter laufen.
2. Lokale Workflows sind schwerer zu beobachten.
3. Auch schwächere Modelle können mit OCR, Datenbanken, Suchindizes und menschlicher Zwischenauswahl effektiv werden.
4. Feinabstimmung und private Retrieval-Speicher erschweren Governance.

Der Bericht behandelt Open Weight daher als Perimeterproblem, nicht als Ursache von Schaden.

Recht und Governance

Die DSGVO widerspricht der Annahme, dass öffentliche personenbezogene Daten frei für jeden Zweck verarbeitet werden dürfen. Sie verlangt Rechtsgrundlage, Zweckbindung, Datenminimierung, Transparenz, Richtigkeit, Speicherbegrenzung, Sicherheit und Betroffenenrechte. EDPB-Leitlinien zum berechtigten Interesse betonen Notwendigkeit und Abwägung. Die Datenschutzkonferenz hebt bei KI-Anwendungen rechtmäßige, zweckgebundene und risikosensible Verarbeitung hervor.

Der EU AI Act ergänzt Regeln zu verbotenen Praktiken, Hochrisikosystemen, Transparenzpflichten und General-Purpose-AI. Für das private Palantir bleiben Lücken: Viele Schäden entstehen informell, unterhalb formaler automatisierter Entscheidungen, in privaten Kontexten, durch lokale Modelle oder durch soziale Nutzung der Ergebnisse.

Internationale Rechtsordnungen unterscheiden sich. Die Analyse ist europäisch geprägt, aber der Mechanismus ist international relevant: Sichtbarkeit wird für Arbeit, Öffentlichkeit und Teilhabe wichtiger, während KI die Aggregation öffentlicher Spuren erleichtert.

Gegenargumente

“Das sind doch öffentliche Daten.”
Öffentlichkeit ist nicht gleichbedeutend mit Einwilligung in unbegrenzte Aggregation, sensible Inferenz

oder dauerhafte Speicherung. Genau hier setzen praktische Obskurität und kontextuelle Integrität an.

“Das ist nur OSINT.”
Legitime OSINT-Arbeit, Journalismus, Menschenrechtsdokumentation und Cybersicherheit sind wichtig. Der Bericht richtet sich nicht gegen öffentliche Recherche, sondern gegen unverhältnismäßige, private, intransparente personenbezogene Aggregation ohne legitimen Zweck, Schutz von Dritten oder Anfechtbarkeit.

“Es gibt keine Prävalenzdaten.”
Richtig. Deshalb behauptet der Bericht keine verbreitete Gegenwartsrealität. Er formuliert einen Schwellen- und Risikomechanismus und fordert Prävalenzforschung.

“Regulierung löst das.”
Teilweise. Bestehendes Recht hilft, aber informelle Profilierung ist schwer nachweisbar, oft privat und häufig nicht als formale Entscheidung dokumentiert.

“Open-Weight-Modelle werden unfair beschuldigt.”
Nein. Sie werden als Governance-Herausforderung behandelt, nicht als Ursache. Offene Modelle haben legitime Vorteile.

Forschungsagenda

1. **Prävalenzstudien:** Wie häufig nutzen private Akteure KI für Personenrecherche, Stalking, Doxxing, Betrugsvorbereitung, Bewerberscreening oder lokale Konflikte?
2. **Agenten-Evaluationen:** Wie stark senken Agenten Aufwand, Zeit und Fehlerkosten gegenüber klassischer Suche?
3. **Institutionelle Expositionsaudits:** Wie aggregierbar sind alte PDFs, Webseiten, Veranstaltungsprogramme, Vereinsseiten, Schularchive und Amtsblätter?
4. **Betroffenenforschung:** Wie verändert erwartete Aggregierbarkeit Verhalten, Rede, Engagement und Sichtbarkeit?
5. **Vergleichende Regulierung:** Wie adressieren unterschiedliche Rechtsordnungen öffentliche personenbezogene Daten, Datenbroker, Stalking, Doxxing und informelle Profilierung?
6. **Schutzmaßnahmen:** Welche Kombinationen aus Metadatenbereinigung, Retention, noindex, Rollenkontakten, Meldewegen und KI-Provider-Safeguards wirken?

Politische und institutionelle Empfehlungen

- Wiederholte personenbezogene Aggregation als eigenständiges Risikomuster anerkennen.
- Institutionelle Veröffentlichungen auf langfristige Aggregierbarkeit prüfen.
- Unnötige personenbezogene Kontaktdaten vermeiden, wo Rollenkontakte genügen.
- PDF-Metadaten vor Veröffentlichung bereinigen.
- Veröffentlichungs-, Kontext- und Überprüfungsdaten ergänzen.
- Für alte Archive Review- und Löschprozesse schaffen.

- Minderjährigen-, Schul-, Jugend- und Familienkontexte besonders sensibel behandeln.
- Anbieter von KI-Systemen sollten wiederholte personenbezogene Aggregation als Missbrauchssignal verstehen.
- Betroffenenhilfe, digitale Gewaltberatung und Beweissicherung an KI-gestützte Beobachtung anpassen.

Grenzen des Berichts

Der Bericht ist keine Prävalenzstudie und beweist nicht, dass private KI-Dossiers bereits verbreitet sind. Er zeigt einen plausiblen, technisch gestützten und gesellschaftlich relevanten Risikomechanismus.

Die technische Lage verändert sich schnell. Einzelne Produktdetails können veralten; dauerhafter sind Risikokette, Evidenzklassen und Forschungsagenda.

Die rechtliche Analyse ist europäisch und deutsch geprägt. Andere Jurisdiktionen haben andere verfassungsrechtliche, datenschutzrechtliche und plattformregulatorische Kontexte.

Die Pilotstudie ist balanciert, aber auf serielle kommunale Publikationen begrenzt. Sie sollte als methodischer und deskriptiver Pilot gelesen werden, nicht als repräsentative Messung aller lokalen Institutionen.

Schlussfolgerung

Das private Palantir ist kein Beweis dafür, dass allgegenwärtige private Überwachung bereits existiert. Es ist eine Warnung vor einer plausiblen und schnell näher rückenden Risikokategorie: Die Schutzwirkung von Zerstreuung, Aufwand, Kontext und Vergessen wird durch KI-Agenten, multimodale Analyse, Speicher und fallende Kosten geschwächt.

Die politische Kernfrage lautet deshalb nicht nur: “Welche Daten sind öffentlich?” Sie lautet: **Wer darf öffentliche Spuren automatisch verbinden, bewerten, speichern und gegen Menschen verwenden?**

Eine demokratische Antwort muss praktische Obskurität nicht romantisieren, aber sie muss sie verteidigen: durch Datenschutz, technische Reibung, klare Grenzen personenbezogener Aggregation, Schutz für Betroffene digitaler Gewalt, faire Regeln für OSINT und Forschung, die weder dramatisiert noch verharmlost.

Methodenanhang: Suchstrategie und Review-Protokoll

Der Bericht ist eine strukturierte narrative Synthese, keine systematische Review im PRISMA-Sinn. Die Recherche umfasste akademische Suchmaschinen, arXiv, Konferenzpapiere, Anbieter- und Produktdokumentation, EU- und Datenschutzquellen, Lageberichte von Strafverfolgungs- und Zivilgesellschaftsorganisationen sowie gezielte Suche nach bekannten Titeln, Autoren und Institutionen.

Zeitrahmen

Der Bericht berücksichtigt Quellen, die bis zum 12. Juni 2026 verfügbar und geprüft waren. Ältere Grundlagenquellen wurden aufgenommen, wenn sie theoretisch oder rechtlich zentral bleiben.

Suchcluster

Cluster	Beispielhafte Suchbegriffe
KI-Agenten	“AI agents web benchmark”, “computer use agent evaluation”, “tool use large language models”
Datenschutztheorie	“practical obscurity”, “contextual integrity”, “public records aggregation privacy”
Doxxing und Harassment	“doxing detection characterization”, “online harassment personal information disclosure”
Technologiegestützte Gewalt	“intimate partner surveillance”, “technology facilitated abuse”, “cyberstalking”
Beschäftigung und Screening	“social media screening hiring discrimination”, “algorithmic hiring bias”
Datenbroker	“data brokers privacy harms”, “people search services privacy”
Chilling Effects	“online surveillance chilling effects”, “digital dataveillance chilling effects”
Governance	“GDPR public personal data legitimate interest”, “OSINT ethics privacy”, “AI Act profiling”

Einschlusskriterien

Quellen wurden aufgenommen, wenn sie technische Fähigkeiten belegen, ein relevantes Schadensfeld dokumentieren, rechtliche oder theoretische Grundlagen liefern, Evidenzklassen schärfen oder konkrete Governance- und Forschungsfragen unterstützen.

Quellen wurden als zentrale Evidenz vermieden, wenn sie rein anekdotisch, promotional ohne technische Substanz, missbrauchsoperativ, redundant oder zu eng jurisdiktionsspezifisch waren.

Pilot-Audit-Protokoll

Die Pilotstudie nutzte einen begrenzten Korpus öffentlicher kommunaler PDF-Publikationen, der im PDF-Datenanhang dokumentiert ist. Das Protokoll:

1. nutzt eine menschlich kuratierte Liste öffentlicher institutioneller PDF-URLs;
2. ruft jedes Dokument nur zur Verarbeitung ab;
3. extrahiert Text und PDF-Metadaten;
4. berechnet dokumentbezogene Aggregatindikatoren;
5. veröffentlicht nur verarbeitete Metriken, Zusammenfassungen und Tabellen;
6. vermeidet Speicherung von Roh-PDFs, Rohtexten, Namen oder Personenprofilen als Publikationsmaterial.

Die Indikatoren sind konservativ und nicht-identifizierend. Namensähnliche Kandidaten werden als rauschbehaftete Expositionsproxys behandelt, nicht als verifizierte Namen.

Quellen

Wissenschaftliche und theoretische Literatur

Acquisti, A. & Fong, C. M. (2020). An experiment in hiring discrimination via online social networks. *Management Science*, 66(3), 1005-1024. https://doi.org/10.1287/mnsc.2018.3269

Brown, C. & Hegarty, K. (2024). Fear and distress: How can we measure the impact of technology-facilitated abuse in relationships? *Social Sciences*, 13(1), 71. https://doi.org/10.3390/socsci13010071

Büchi, M., Festic, N. & Latzer, M. (2022). The chilling effects of digital dataveillance: A theoretical model and an empirical research agenda. *Big Data & Society*, 9(1). https://doi.org/10.1177/20539517211065368

Crain, M. (2021). The untamed and discreet role of data brokers in surveillance capitalism: A transnational and interdisciplinary overview. *Internet Policy Review*, 10(2). https://policyreview.info/articles/analysis/untamed-and-discreet-role-data-brokers-surveillance-capitalism-transnational-and

Hartzog, W. & Stutzman, F. (2010). The case for online obscurity. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1597745

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79(1), 119-158. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/

Penney, J. W. (2016). Chilling effects: Online surveillance and Wikipedia use. *Berkeley Technology Law Journal*, 31(1), 117-182. https://btlj.org/data/articles2016/vol31/31_1/0117_0182_Penney_ChillingEffects_WEB.pdf

Snyder, P., Doerfler, P., Kanich, C. & McCoy, D. (2017). Fifteen minutes of unwanted fame: Detecting and characterizing doxing. In *Proceedings of the 2017 Internet Measurement Conference* (pp. 432-444). ACM. https://doi.org/10.1145/3131365.3131385

Tseng, E. et al. (2020). The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums. In *29th USENIX Security Symposium*. USENIX Association. https://www.usenix.org/conference/usenixsecurity20/presentation/tseng

KI-Agenten, Modelle und Benchmarks

Anthropic. (2026). *Computer use tool*. https://docs.anthropic.com/en/docs/build-with-claude/computer-use

Anthropic. (2026). *Tool use*. https://docs.anthropic.com/en/docs/build-with-claude/tool-use/overview

Kapoor, S. et al. (2024). *AI agents that matter*. arXiv. https://arxiv.org/abs/2407.01502

Mistral AI. (2025). *Le Chat dives deep*. https://mistral.ai/news/le-chat-dives-deep/

Microsoft. (2025). *Microsoft Build 2025: The age of AI agents and building the open agentic web*. https://blogs.microsoft.com/blog/2025/05/19/microsoft-build-2025-the-age-of-ai-agents-and-building-the-open-agentic-web/

OpenAI. (2025). *Computer-using agent*. https://openai.com/index/computer-using-agent/

OpenAI. (2025). *Introducing deep research*. https://openai.com/index/introducing-deep-research/

OpenAI. (2026). *Introducing ChatGPT agent*. https://openai.com/index/introducing-chatgpt-agent/

OpenAI. (2026). *Introducing GPT-5*. https://openai.com/index/introducing-gpt-5/

Xie, T. et al. (2024). *OSWorld: Benchmarking multimodal agents for open-ended tasks in real computer environments*. arXiv. https://arxiv.org/abs/2404.07972

Zhou, S. et al. (2023). *WebArena: A realistic web environment for building autonomous agents*. arXiv. https://arxiv.org/abs/2307.13854

Online-Mind2Web authors. (2025). *An illusion of progress? Assessing the current state of web agents*. arXiv. https://arxiv.org/abs/2504.01382

Open Weight, Kosten und Modellverbreitung

DeepSeek. (2025). *DeepSeek-R1 release*. https://api-docs.deepseek.com/news/news250120

Hugging Face. (2025). *DeepSeek-R1 model card*. https://huggingface.co/deepseek-ai/DeepSeek-R1

Meta AI. (2025). *The Llama 4 herd*. https://ai.meta.com/blog/llama-4-multimodal-intelligence/

OECD. (2024). *AI openness: A primer for policymakers*. OECD Publishing. https://www.oecd.org/en/publications/ai-openness_02f73362-en.html

Qwen. (2025). *Qwen3-Coder*. https://qwenlm.github.io/blog/qwen3-coder/

QwenLM. (2025). *Qwen3-Coder GitHub repository*. https://github.com/QwenLM/Qwen3-Coder

Stanford Institute for Human-Centered Artificial Intelligence. (2025). *AI Index Report 2025*. https://hai.stanford.edu/ai-index/2025-ai-index-report

Stanford Institute for Human-Centered Artificial Intelligence. (2026). *AI Index Report 2026*. https://hai.stanford.edu/ai-index/2026-ai-index-report

Datenschutz, digitale Gewalt und Cybercrime

Bundeskriminalamt. (2026). *Cybercrime-Lagebilder*. https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/Lagebilder/Cybercrime/cybercrime_node.html

Europol. (2025). *Internet Organised Crime Threat Assessment 2025*. https://www.europol.europa.eu/cms/sites/default/files/documents/Steal-deal-repeat-IOCTA_2025.pdf

Europol. (2025). *Serious and Organised Crime Threat Assessment 2025*. https://www.europol.europa.eu/cms/sites/default/files/documents/EU-SOCTA-2025.pdf

Federal Bureau of Investigation Internet Crime Complaint Center. (2026). *2025 IC3 Annual Report*. https://www.ic3.gov/AnnualReport/Reports/2025_IC3Report.pdf

HateAid & Technische Universität München. (2025). *Angegriffen und alleingelassen: Wie sich digitale Gewalt auf politisches Engagement auswirkt*. https://hateaid.org/wp-content/uploads/2025/01/hateaid-tum-studie-angegriffen-und-alleingelassen-2025.pdf

New America. (2025). *Preserving privacy: An impact framework for OSINT and AI*. https://www.newamerica.org/insights/preserving-privacy-an-impact-framework/

Privacy International. (2026). *Challenge against Clearview AI in Europe*. https://privacyinternational.org/legal-action/challenge-against-clearview-ai-europe

Recht, Regulierung und Standards

Cornell Legal Information Institute. (n.d.). *U.S. Department of Justice v. Reporters Committee for Freedom of the Press, 489 U.S. 749*. https://www.law.cornell.edu/supremecourt/text/489/749

European Commission. (2026). *Regulatory framework on artificial intelligence*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Data Protection Board. (2024). *Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR*. https://www.edpb.europa.eu/system/files/2024-10/edpb_guidelines_202401_legitimateinterest_en.pdf

European Union. (2016). *Regulation (EU) 2016/679: General Data Protection Regulation*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Datenschutzkonferenz. (2024). *Orientierungshilfe KI und Datenschutz*. https://www.datenschutzkonferenz-online.de/media/oh/20240506_DSK_Orientierungshilfe_KI_und_Datenschutz.pdf

Bundesministerium der Justiz. (2026). *Gesetz gegen digitale Gewalt*. https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/2026_Gesetz_gegen_digitale_Gewalt.html

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework*. https://www.nist.gov/itl/ai-risk-management-framework

---

Impressum und Autorennotiz

Autor und redaktionelle Verantwortung: Benjamin Metzig, Wissenschaftswelle.de. Wissenschaftswelle.de ist ein unabhängiges redaktionelles Wissensprojekt zu Wissenschaft, Gesellschaft, Technologie, Kultur und öffentlichem Verständnis von Evidenz.

Projektwebsite: https://www.wissenschaftswelle.de/

Empfohlene Zitierweise: Metzig, B. (2026). *Das private Palantir: Forschungsbericht über KI-Agenten, öffentliche Daten und das Ende praktischer Obskurität*. Wissenschaftswelle.de.

Persistenter Identifikator: Noch kein DOI vergeben. Der Bericht ist für spätere Repository-Ablage, DOI-Registrierung und versionierte Zitierung strukturiert.

Redaktioneller Hinweis: Digitale Forschungs- und Schreibwerkzeuge wurden bei Strukturierung, Sprachfassung und Konsistenzprüfung unterstützend genutzt. Sie wurden nicht als Autorität für Tatsachenbehauptungen verwendet. Quellenprüfung, Interpretation, Schlussbewertung und Veröffentlichungsverantwortung liegen bei Benjamin Metzig.

Datenverfügbarkeit: Für eine reine PDF-Veröffentlichung wird die Pilotstudie durch einen PDF-Datenanhang begleitet. Dieser enthält öffentliche Quellen-URLs, Korpuszusammenfassung, stadtbezogene Ergebnisse und dokumentbezogene Aggregatindikatoren. Roh-PDFs, Rohtexte, Namen und personenbezogene Datensätze sind nicht Teil des Publikationsmaterials.

Codeverfügbarkeit: Das Publikationspaket erfordert keinen lokalen Codezugang. Der Bericht beschreibt Korpuslogik, Indikatoren, Score-Komponenten und ethische Schutzmaßnahmen. Rechnerische Hilfen wurden zur URL-Validierung und Aggregatzählung genutzt, unter der Bedingung, dass keine Rohtexte, Namen oder Personenprofile als Publikationsoutput erhalten bleiben.

Interessenkonflikte: Benjamin Metzig ist Gründer, Autor und redaktionell verantwortlicher Herausgeber von Wissenschaftswelle.de. Es werden keine finanziellen Interessenkonflikte erklärt.