



365 Bericht von Hannes Ullrich, Jonas Hannane, Tomaso Duso, Christian Peukert und Luis Aguiar

Daten als Wettbewerbsvorteil: Regulierer und Wettbewerbsbehörden sollten aufmerksam bleiben

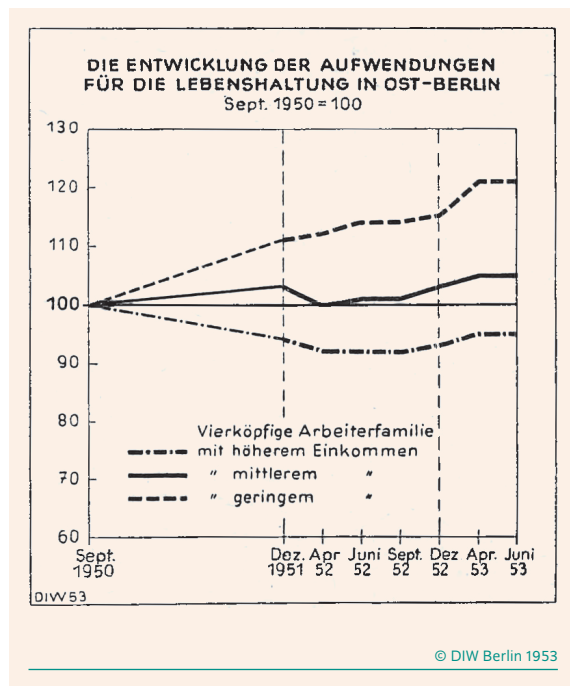
- Digitalunternehmen profitieren von Datenmengen – auch wenn der zusätzliche Nutzen weiterer Daten abnimmt
- Google hat durch die Mengen an Daten, die es sammelt, einen Wettbewerbsvorteil
- Regulierung kann nötig sein, um zukünftige Markteintrittsschranken zu verhindern

374 Interview mit Hannes Ullrich

378 Kommentar von Claudia Kemfert und Christian von Hirschhausen

Bundesregierung überschätzt Bedarf an Flüssiggas-Infrastruktur

Die Entwicklung der Lebenshaltungskosten im sowjetischen Besatzungsgebiet



Zu den grundlegenden Bestimmungen für die neuen wirtschaftspolitischen Maßnahmen in der Sowjetzone zählt der „Beschluss über die Grundsätze der Preispolitik“. Das wichtigste Ziel dieses „Beschlusses“ ist die Schaffung eines einheitlichen Preisniveaus, das bereits seit 1949 angestrebt wird. Welche Wirkung sich hieraus auf die Entwicklung der Lebenshaltungskosten ergibt, zeigt die folgende Graphik.

Aus dem Wochenbericht Nr. 27 vom 3. Juli 1953

IMPRESSUM



DIW Berlin — Deutsches Institut für Wirtschaftsforschung e. V.

Mohrenstraße 58, 10117 Berlin

www.diw.de

Telefon: +49 30 897 89-0 Fax: -200

90. Jahrgang 5. Juli 2023

Herausgeberinnen und Herausgeber

Prof. Dr. Tomaso Duso; Sabine Fiedler; Prof. Marcel Fratzscher, Ph.D.;
Prof. Dr. Peter Haan; Prof. Dr. Claudia Kemfert; Prof. Dr. Alexander S. Kritikos;
Prof. Dr. Alexander Kriwoluzky; Prof. Dr. Lukas Menkhoff; Prof. Karsten
Neuhoff, Ph.D.; Prof. Dr. Carsten Schröder; Prof. Dr. Katharina Wrohlich

Chefredaktion

Prof. Dr. Pio Baake; Claudia Cohnen-Beck; Sebastian Kollmann;
Kristina van Deuverden

Lektorat

Dr. Daniel Graeber

Redaktion

Rebecca Buhner; Dr. Hella Engerer; Ulrike Fokken; Petra Jasper; Kevin Kunze;
Sandra Tubik

Vertrieb

DIW Berlin Leserservice, Postfach 74, 77649 Offenburg

leserservice@diw.de

Telefon: +49 781 639 67 20

Gestaltung

Roman Wilhelm, Stefanie Reeg, Eva Kretschmer, DIW Berlin

Umschlagmotiv

© imageBROKER / Steffen Diemer

Satz

Satz-Rechen-Zentrum Hartmann + Heenemann GmbH & Co. KG, Berlin

Druck

USE gGmbH, Berlin

ISSN 0012-1304; ISSN 1860-8787 (online)

Nachdruck und sonstige Verbreitung – auch auszugsweise – nur mit
Quellenangabe und unter Zusendung eines Belegexemplars an den
Kundenservice des DIW Berlin zulässig (kundenservice@diw.de).

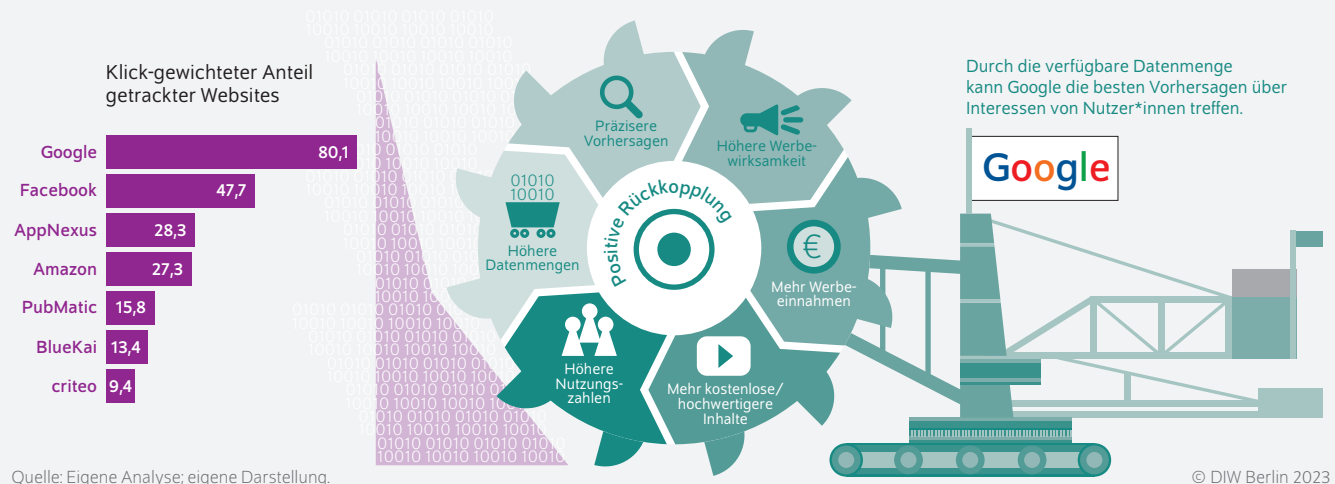
Abonnieren Sie auch unseren DIW- und/oder Wochenbericht-Newsletter
unter www.diw.de/newsletter

Daten als Wettbewerbsvorteil: Regulierer und Wettbewerbsbehörden sollten aufmerksam bleiben

Von Hannes Ullrich, Jonas Hannane, Tomaso Duso, Christian Peukert und Luis Aguiar

- Internetnutzer*innen können automatisiert beobachtet werden, um Profile für zielgerichtete Werbung zu erstellen
- Digitalunternehmen profitieren von Datenmengen – auch wenn der zusätzliche Nutzen weiterer Daten abnimmt
- Möglicher Netzwerkeffekt: Mehr Daten führen zu präziserer Personalisierung von Inhalten und Werbung, diese führen zu mehr Nutzer*innen, von denen sich zusätzliche Daten ergeben
- Google-Konzern Alphabet kann durch die Mengen an Daten, die er sammelt, einen Wettbewerbsvorteil erlangen
- Regulierung kann nötig sein, um zukünftige Markteintrittsschranken zu verhindern

Onlinekonzerne profitieren von Daten, aber Google sticht heraus: Daten könnten dem Google-Betreiber Alphabet eine uneinholbare Position im Markt beschern



ZITAT

„Seinen Vorsprung konnte Google durch seine Pionierposition bekommen, und auch, weil Behörden spät erkannt haben, dass Daten ein zentraler Wettbewerbsfaktor sind. Mit Blick auf neue datenbasierte Technologien wie ChatGPT liegt nahe: Frühzeitige Regulierung kann helfen, den Markt offen für neue Angebote zu halten.“

— Hannes Ullrich —

MEDIATHEK



Audio-Interview mit Hannes Ullrich
www.diw.de/mediathek

Daten als Wettbewerbsvorteil: Regulierer und Wettbewerbsbehörden sollten aufmerksam bleiben

Von Hannes Ullrich, Jonas Hannane, Tomaso Duso, Christian Peukert und Luis Aguiar

ABSTRACT

Das umfangreiche Sammeln persönlicher Daten im Internet hat in den letzten zwei Jahrzehnten einen großen Markt für personalisierte Inhalte ermöglicht. Tausende kleine und große Unternehmen arbeiten an der Sammlung und Zusammenführung von Daten über das Nutzer*innenverhalten auf Websites, um über Eigenschaften und Interessen von Nutzer*innen zu lernen. Ganz an der Spitze befinden sich digitale Unternehmen wie Alphabet (Google), Amazon und Meta (Facebook), die den Großteil ihrer Umsätze durch Online-Werbung erwirtschaften. Diese Studie zeigt empirisch, dass eine steigende Menge an Daten zu einer besseren Vorhersagequalität von Personenmerkmalen führt, die für personalisierte Inhalte essenziell sind. Dennoch nehmen die Verbesserungen der Vorhersagequalität ab. Je mehr Daten dazukommen, desto weniger zusätzlichen Nutzen haben sie. Man kann hier von abnehmenden Skalenerträgen sprechen. Google, das bisher Klick-gewichtet auf über 80 Prozent der gängigsten Websites Nutzer*innenverhalten beobachten konnte, erreichte, bei gleichzeitig steigenden Nutzer*innenzahlen und beobachteten Websitebesuchen, jedoch signifikant weniger stark abnehmende Skalenerträge als seine Wettbewerber. Wenn aber Google allein von weiteren Daten durchweg mehr profitieren kann als seine Wettbewerber, können selbst innovative kleinere Unternehmen mangels Daten nur geringe Chancen haben, konkurrenzfähige Produkte zu vermarkten. Es würde sich eine uneinholbare Marktposition für den Google-Konzern ergeben, was zu Nachteilen für Konsument*innen führt.

Daten sind ein zentraler Wettbewerbsfaktor in digitalen Märkten. Sie ermöglichen innovative Dienstleistungen und Produkte von großem Wert für Konsument*innen. Neben Erfolgsmodellen wie der Google-Suchmaschine oder Amazons Online-Shop haben zuletzt große datenbasierte Sprachmodelle (Large Language Models, LLM) viel Aufmerksamkeit erregt. Sie werden zum Beispiel in ChatGPT vom Unternehmen OpenAI und auch in Produkten von Microsoft angewendet. Die große Bedeutung von Datenmengen wird jedoch zu einem gesellschaftlichen und wettbewerbsrechtlichen Problem, wenn der Zugang zu Daten einzelnen Unternehmen einen so großen Qualitäts- oder Kostenvorsprung verschafft, dass es zu Monopolisierung und Marktzutrittsschranken kommt, wodurch Märkte für neue Wettbewerber verschlossen werden.

Das kann bedeuten, dass nicht nur andere Unternehmen, sondern auch Konsument*innen Nachteile haben. Wenn Märkte hohe Zutrittsschranken haben, bleiben sie potenziell für Innovationen unzugänglich. Für Endnutzer*innen kann dies bedeuten, dass ihnen mittelfristig etwa neue technische Möglichkeiten, Verbesserungen in der Nutzungsfreundlichkeit oder gänzlich neue Online-Dienste, die sich zu besseren Alternativen bestehender Dienste entwickeln können, vorenthalten bleiben.

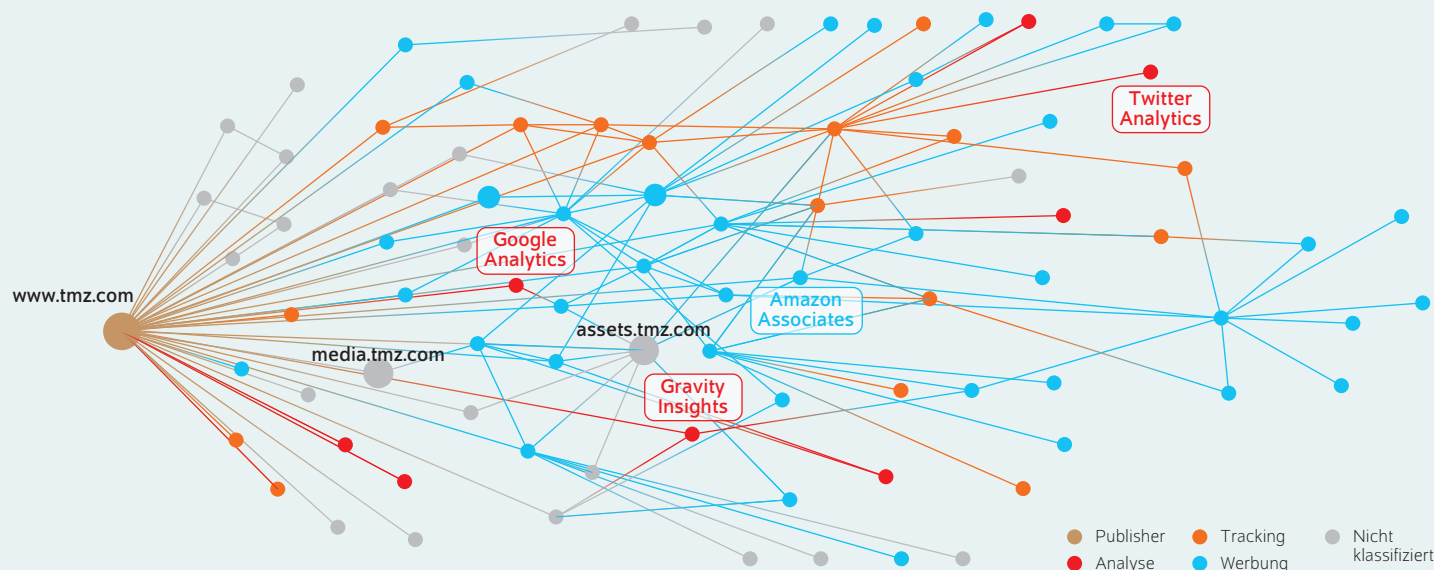
In der Tat deuten mehrere theoretische Studien darauf hin, dass Unternehmen in digitalen Märkten durch ihre Datenmengen Kippunkte erreichen können, aus denen unüberwindbare Barrieren zum Markteintritt für andere Unternehmen folgen – selbst, wenn es anfangs nur kleine Unterschiede zwischen den Unternehmen gab.¹ Deshalb drehen sich viele politische Debatten darum, wie der Zugang zu Daten in Plattformmärkten reguliert werden kann, zum Beispiel im Digital Markets Act (DMA) der Europäischen Union.²

¹ Vgl. Jan Krämer und Daniel Schnurr (2022): Big Data and Digital Markets Contestability: Theory of Harm and Data Access Remedies. *Journal of Competition Law & Economics* 18(2), 255–322. Jens Prüfer und Christoph Schottmüller (2021): Competing with Big Data. *Journal of Industrial Economics* 69, 967–1008. Bundeskartellamt (2017): Big Data und Wettbewerb (online verfügbar, abgerufen am 16. Juni 2023). Dies gilt auch für alle anderen Online-Quellen dieses Berichts, sofern nicht anders vermerkt.

² Vgl. Digital Markets Act der Europäischen Union (online verfügbar).

Abbildung 1

Tracker-Netzwerk am Beispiel der Website eines US-Boulevardmediums



Quelle: Eigene Analyse; eigene Darstellung.

© DIW Berlin 2023

Verschiedene Unternehmen sammeln Daten auf Websites durch Produkte mit unterschiedlichen Zwecken. Diese Daten werden an andere Unternehmen weitergegeben oder verkauft.

In der Studie, auf der dieser Wochenbericht basiert, wird überprüft, ob das großflächige Sammeln von Daten über das allgemeine Surfverhalten von Internetnutzer*innen den größten Online-Werbepattformen einen unüberwindbaren Wettbewerbsvorsprung verleihen kann.³

Daten sind das wirtschaftliche Fundament von Tech-Giganten

Bekannte Unternehmen wie Alphabet (Anbieter von Google Search, Google Maps, Google Mail, Google Analytics, Google Play, Android und zahlreicher weiterer Produkte), Meta (Facebook, Instagram, WhatsApp), Microsoft (Windows, Office, Bing, Azure) und Amazon (Einzelhandel, Amazon Web Services, Amazon Ads) betreiben eigene Ökosysteme von vielen unterschiedlichen Produkten und Plattformen, in welchen Nutzungsdaten generiert werden.⁴ Ihr Geschäftsmodell fußt unter anderem darauf, diese Daten auf unterschiedliche Arten und Weisen zu verknüpfen, zu nutzen und zu monetarisieren.

Daten sind wertvoll für Unternehmen, denn sie ermöglichen eine Verbesserung von digitalen Inhalten, zum Beispiel anhand von Empfehlungs-Algorithmen für Video- und

Audiomedien sowie vielen anderen Online-Diensten, aber auch von Softwareprodukten und mehr. So können digitale Plattformen vielseitige datenbasierte Produkte anbieten, von denen Konsument*innen zunächst profitieren.

Online-Werbung ist das Paradebeispiel eines Produkts, bei dem Daten den wichtigsten Input darstellen. Für einige der größten digitalen Unternehmen wie Alphabet, Meta und zunehmend auch Amazon ist dieses Produkt von zentraler Bedeutung. Ihre Gewinne stammen zu einem großen Teil aus dem Verkauf von Werbung auf ihren Online-Plattformen wie Instagram, Facebook und Google Search oder auf Websites, die sie nicht selbst betreiben.⁵ Die Werbeeinnahmen von Alphabet und Meta zusammen beliefen sich im Jahr 2022 auf 338 Milliarden US-Dollar.⁶ Dieser Erfolg wird unter anderem der Fähigkeit dieser Unternehmen zugeschrieben, so zielgenau Werbung zu schalten, dass sie bei Konsument*innen häufiger zum Klick oder direkt zum Kauf führt.

Datenmengen können bestimmend sein

Gezielte Werbung basiert auf umfangreichen Daten, die von Unternehmen gesammelt und analysiert werden. Diese Daten entstehen häufig als Nebenprodukt bei der Nutzung von digitalen Anwendungen oder beim Surfen im Internet. Hierbei können zum Beispiel Daten über das verwendete

³ Vgl. Luis Aguiar et al. (2023): Returns to Web Tracking Data, mimeo (online verfügbar).

⁴ In der restlichen Publikation wird der Markenname Google statt der Name des Trägerkonzerns Alphabet benutzt, da es zu großen Teilen um Google-Dienste geht und dies die deutlich bekanntere Bezeichnung ist. Weiterhin wird im Zusammenhang mit der Untersuchung Facebook genannt, da die Stichprobe aus der Zeit stammt, in der das Unternehmen noch nicht Meta hieß.

⁵ Vgl. Competition and Markets Authority (2020): Online platforms and digital advertising market study (online verfügbar).

⁶ Vgl. Darstellung von Statista (online verfügbar).

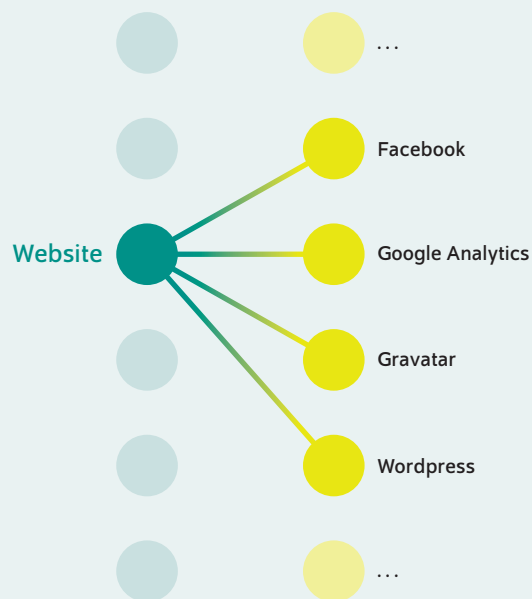
Abbildung 2

Verschiedene Tracker im HTML-Code einer Website

```

...
<div id="rr social widgets facebook">
  <iframe src="//www.facebook.com/plugins/
    like.php?href=http%3A%2Fwww.facebook.com
    %2Ftechcrunch&layout=standard..." ...
  </iframe>
</div>
...
<script type="text/javascript">
  var gaJsHost = (("https:" == document.
    location.protocol)? "https://ssl." : "http://www.");
  document.write(unescape("%3Cscript
    src='" + gaJsHost +
    "google-analytics.com/ga.js'
    type='text/javascript'%3E%3C/script%3E"));
</script>
...
<script src='http://s.gravatar.com/js/
  type='text/javascript'
  gprofiles.js?aa#038;ver=3.5-alpha-21304'>
</script>
...
<noscript>
  
</noscript>
...

```



Quelle: Eigene Analyse; eigene Darstellung.

© DIW Berlin 2023

Websites laden zahlreiche Tracker verschiedener Anbieter, die von den Seitenbetreibern in den Quellcode eingebaut werden.

Gerät, insbesondere aber auch die Adressen einzelner besuchter Websites erhoben werden. Neben bekannten Unternehmen wie Alphabet und Meta sammeln mehr als 4000 eher unbekannte Datenbroker ebendiese Daten (Abbildung 1). Diese Unternehmen konkurrieren im Internet, um mit Analyse-, Technologie- oder anderen Dienstleistungen auf möglichst vielen und relevanten Websites integriert zu werden. Sie bieten also Leistungen, die für Websitebetreiber Vorteile bieten. Teil dieses Handels ist dann, dass die Unternehmen dort die Klicks von Nutzer*innen beobachten und, auf dem Klickverhalten basierend, personenspezifische Profile erstellen können.⁷ Demografische Profile, zum Beispiel eine Unterteilung von Personen nach Alter, Geschlecht oder Wohnort, werden dann verwendet, um individuelle Interessen als Grundlage für personalisierte Werbung abzuleiten.⁸ Werbeanzeigen können so für gewünschte Zielgruppen verkauft werden.

Das Ergebnis aus Sicht der Nutzer*innen ist aus der Internetsuche bekannt. Wenn man bei Google Search oder Bing nach „Herrentasche“ sucht, ist die Wahrscheinlichkeit hoch, dass in der Ergebnisliste eine Werbeanzeige für eine Handtaschenmarke erscheint. Welche Herrentaschen man zu sehen

bekommt, kann wiederum von der Art des eigenen Profils abhängen. Diese Profile werden erstellt, indem Google, Meta und andere Datenbroker auf vielen Websites kleine Code-Schnipsel platzieren (Abbildung 2), die Informationen über die Nutzer*innen – zum Beispiel von welcher Website sie kommen, wie lange sie auf der Seite bleiben, welche Links sie anklicken – an die Server der trackenden Unternehmen senden können. Durch das Speichern von sogenannten Cookies auf den lokalen Rechnern von Nutzer*innen können diese über Websites hinweg identifiziert werden.⁹ Anhand dieser Informationen können Unternehmen die Charakteristika der Nutzer*innen abschätzen. Zum Beispiel wird ein Nutzer, der häufig die Website eines Sportmagazins besucht, das auf eine junge Kundschaft spezialisiert ist, bei der Suche nach dem Begriff „Herrentasche“ eher ein Produkt für 20-Jährige als für 60-Jährige finden.

Datennetzwerkeffekte machen Profile präziser

Um auf dem Markt für personalisierte Werbung zu bestehen, können zum einen ausreichende Datenmengen und zum anderen technologische Innovationen in der Verwertung der Daten notwendig sein.

⁷ Vgl. Hannes Ullrich et al. (2022): Plattformen wie Facebook können mehr als die Hälfte der Internetaktivität beobachten. DIW Wochenbericht Nr. 29/30, 400–406 (online verfügbar).

⁸ Vgl. Nico Neumann, Catherine E. Tucker und Timothy Whitfield (2019): Frontiers: How effective is third-party consumer profiling? Evidence from field studies. Marketing Science 38 (6), 918–926. Jon Keegan und Joel Eastwood (2023): From "Heavy Purchasers" of Pregnancy Tests to the Depression-Prone: We Found 650,000 Ways Advertisers Label You. The Markup, 8. Juni 2023 (online verfügbar).

⁹ Bei First-Party-Cookies geschieht dies durch den Anbieter der besuchten Website. Werden diese von Drittanbietern geladen und Daten außerhalb der besuchten Website verknüpft, so spricht man von Third-Party-Cookies. Die verpflichtende Einholung der Zustimmung zur Nutzung von Third-Party-Cookies ist ein Hauptpunkt der seit Mai 2018 geltenden EU-Datenschutz-Grundverordnung (EU-DSGVO).

Die Profile werden von Datenbrokern und Plattformen mit Methoden des maschinellen Lernens erstellt. So werden aus den von ihnen gesammelten Daten Vorhersagen über die Merkmale der Profilinhaber*innen getroffen. Eine Besonderheit in der Auswertung von Online-Nutzungsdaten sind sogenannte Datennetzwerkeffekte, welche eine sehr hohe Attraktivität und Qualität von digitalen Diensten und Produkten ermöglichen können.

Datennetzwerkeffekte entstehen durch eine positive Rückkopplung: Präzisere Vorhersagen ermöglichen ein besseres Erreichen der Zielgruppe durch Werbetreibende. Dies erhöht die Wirksamkeit der Werbung, wodurch im Gegenzug Online-Plattformen, auf denen die Werbung geschaltet wird, höhere Preise für die Bereitstellung der Werbefläche erzielen. Die höheren Einnahmen der Plattformen ermöglichen diesen, mehr kostenlose oder hochwertigere Inhalte bereitzustellen. Dies lockt wiederum mehr Nutzer*innen auf die Plattform, die so mehr Daten erhält. Die höheren Datenmengen können wiederum die Präzision der Vorhersagen verbessern, wodurch sich der Zyklus fortsetzt (Abbildung 3).

In zweiseitigen Märkten, in denen sich zwei Kund*innen-gruppen in ihrem Nachfrageverhalten gegenseitig beeinflussen, etwa Internetnutzer*innen und Werbetreibende, ist häufig eine Winner-takes-all-Dynamik zu beobachten, welche zum Kippen dieser Märkte führt. Genauso können auch Datennetzwerkeffekte zu Monopolisierungstendenzen beitragen.

Zusätzliche Daten ermöglichen zielgenauere Werbung, der zusätzliche Ertrag nimmt aber ab

Die Relevanz solcher Datennetzwerkeffekte hängt vom Wert wachsender Datenmengen für die Produktqualität und damit für den Unternehmenserfolg im Markt ab. Es liegt auf der Hand, dass die Genauigkeit der Vorhersagen zunimmt, je mehr Daten einem trackenden Unternehmen zur Verfügung stehen: Daten weisen sogenannte Skalenerträge auf (Kasten 1). Für einzelne Anwendungen von Internet- und Produktsuchmaschinen wurde in bestehenden Studien gezeigt, dass Skalenerträge von Daten abnehmend sind.¹⁰ Wenn ein trackendes Unternehmen Informationen über weniger Nutzer*innen hat, ist der Wert, eine*n einzelne*n zusätzliche*n Nutzer*in zu beobachten, groß. Der Wert zusätzlicher Daten ist aber viel kleiner, wenn schon Daten über Millionen von Nutzer*innen vorhanden sind. Das hat grundlegende Auswirkungen auf den Wettbewerb und die Entwicklung eines Marktes.

¹⁰ Vgl. Di He et al. (2017): Scale effects in web search. In Web and Internet Economics, WINE 2017 Proceedings, 294–310. Patrick Bajari et al. (2019): The impact of big data on firm performance: An empirical investigation. AEA Papers & Proceedings 109, 33–37. Tobias Klein et al. (2023): How Important Are User-generated Data For Search Result Quality? Experimental Evidence. CEPR Discussion Paper DP17934. Christian Peukert, Ananya Sen und Jörg Claussen (2023): The Editor and the Algorithm: Recommendation Technology in Online News. Im Erscheinen bei Management Science. Eine Studie findet lokal zunehmende Skalenerträge von Daten für die Qualität einer Suchmaschine und diskutiert die damit einhergehenden Konsequenzen für den Wettbewerb. Vgl. Maximilian Schaefer und Geza Sapi (2023): Complementarities in Learning from Data: Insights from General Search, mimeo (online verfügbar).

Abbildung 3

Datennetzwerkeffekte



Quelle: Eigene Darstellung.

© DIW Berlin 2023

Mehr Daten können zu mehr Einnahmen, mehr Nutzer*innen und daraus folgend zu noch mehr Daten führen.

Lässt sich schon mit sehr geringen Datenmengen eine hohe Vorhersagekraft erreichen, so ist es auch für kleine Konkurrenten mit wenigen Daten leicht, ein erfolgreiches Produkt anzubieten. Unterscheidet sich aber der Wert zusätzlicher Daten systematisch zwischen konkurrierenden Firmen, so können Datennetzwerkeffekte zu problematischen Marktstrukturen führen.

Durch den Vergleich von Skalenerträgen über viele Unternehmen hinweg kann in der Studie analysiert werden, ob die Verfügbarkeit von Daten zu einer möglichen Markteintrittsbarriere für konkurrierende Unternehmen werden kann. Das wäre dann der Fall, wenn die Skalenerträge eines Unternehmens, das Tracker auf den meisten Websites installiert hat und daher viele und vielfältige Daten sammelt, in Bezug auf die Anzahl der Nutzer*innen und Websites weniger stark abnehmen als für andere, kleinere Konkurrenten.

Um diese Fragestellung zu beantworten, kommen Methoden des maschinellen Lernens zum Einsatz, die es ermöglichen, das Nutzungsverhalten im Internet, beispielsweise regelmäßige Besuche auf bestimmten Websites, mit persönlichen Eigenschaften wie dem Alter oder Geschlecht der Nutzer*innen statistisch zu verknüpfen.

Kasten 1

Skalenerträge für datenbasiertes Lernen

Die zentrale Frage der Studie lautet, wie sehr sich die Genauigkeit der von Datenbrokern und anderen Unternehmen, auf Vorhersagen basierten, erstellten Konsument*innenprofile verändert, je nachdem wieviel Daten zur Verfügung stehen. In der Volkswirtschaftslehre spricht man von sogenannten Skalenerträgen, um die Abhängigkeit einer Produktionsmenge, in dieser Studie also die Vorhersagequalität, von der Menge eingesetzter Produktionsfaktoren, das sind hier die Daten, zu beschreiben. Dabei unterscheidet man zwischen konstanten, zunehmenden oder abnehmenden Skalenerträgen, je nachdem ob die Produktionsmenge bei einer proportionalen Erhöhung der Produktionsfaktoren um einen Faktor μ ebenfalls um μ ansteigt, mehr als μ ansteigt oder weniger als μ ansteigt (Abbildung).

Eine Quantifizierung von datenbasierten Skalenerträgen ist essenziell, um ein besseres Verständnis über die Dynamiken in zweiseitigen Märkten zu erhalten. Aufgrund der durch Netzwerkeffekte generierten positiven Rückkopplung können nicht- oder nur schwach abnehmende Skalenerträge zu Markteintrittsbarrieren und Monopolisierung von Märkten führen. Dadurch wurde ein Eingreifen von Wettbewerbsbehörden oft motiviert.

Daten über Internetverläufe können entlang von zwei Dimensionen gesammelt werden: Es können sowohl mehr Daten über mehr Internetnutzer*innen gesammelt werden als auch mehr Daten über die einzelnen Internetnutzer*innen. Die Studie analysiert die Skalenerträge aus Daten und zeigt, wie sich die Vorhersagequalität verbessert, wenn mehr Daten in jeder Dimension für sich genommen verfügbar sind und ob es eine Wechselwirkung zwischen beiden Dimensionen gibt.

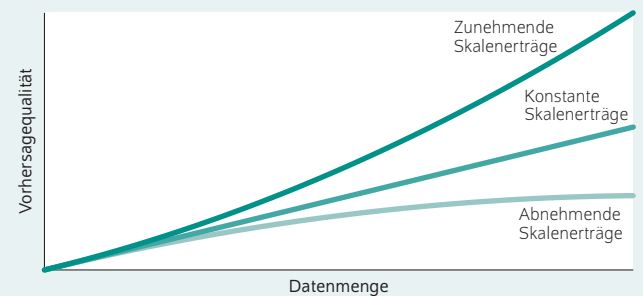
Grundlage der Analyse sind Daten einer repräsentativen Stichprobe der Marktforschungsfirma Nielsen aus dem Jahr 2016. In einer Erhebung haben 4989 Internetnutzer*innen in den USA persönliche Merkmale angegeben und ihr Surfverhalten über ein Jahr speichern lassen, indem ein Browser-Plugin den gesamten Verlauf aller von den individuellen Nutzer*innen besuchten Websites aufzeichnete. Hierbei wurden sowohl die Anzahl an getätigten Klicks pro Website als auch die Verweildauer auf der entsprechenden Seite aufgenommen. Aus einem zweiten Datensatz der Initiative HTTP Archive wird für jede besuchte Website die Information übernommen, welche externen Dienstleister beim Laden der Website Informationen übertragen konnten.¹¹ Durch das Verknüpfen dieser beiden Datenquellen kann der durch individuelle externe Dienstleister beobachtbare Anteil des Surfverhaltens für alle Internetnutzer*innen quantifiziert werden.

Basierend auf den jeweiligen Daten zu besuchten Websites, über die jedes einzelne Unternehmen verfügt, wird in der

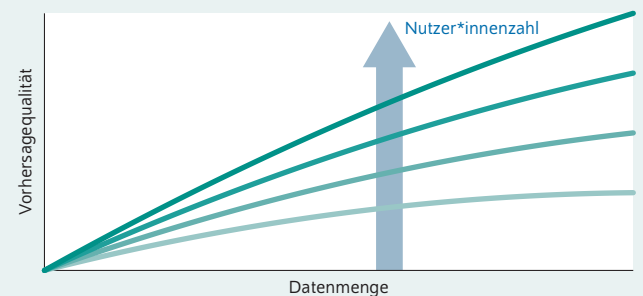
Abbildung

Skalenerträge

Mögliche Formen von Skalenerträgen



Beispielhaft: Unterschiedlich abnehmende Skalenerträge in Abhängigkeit der Anzahl von Nutzer*innen



Quelle: Eigene Darstellung.

© DIW Berlin 2023

Studie ein Algorithmus zur Klassifizierung von Personen nach zehn Eigenschaften angewandt: Alter in vier Kategorien, Geschlecht, hohes Einkommen, hohes Bildungsniveau, Beschäftigungsstatus, Kinder im Haushalt und Lokalisierung in einem wahlentscheidenden US-Bundesstaat. Um den Wert zusätzlicher Daten für die Qualität der Klassifikation zu messen, wird stets derselbe Algorithmus verwendet, während die Menge der Daten variiert (Kasten 2).¹² Das sind die Anzahl der beobachtbaren und besuchten Websites, gewichtet durch die Anzahl an Klicks auf einer Website, sowie die Anzahl an beobachteten Nutzer*innen.

Aus der Analyse wird deutlich, dass die Qualität der Klassifizierung mit zunehmenden Daten steigt. Der Zuwachs der Qualität wird aber immer kleiner. Am Beispiel der Variablen Geschlecht zeigt sich: Sogar mit wenigen Daten, etwa mit nur 20 Prozent der klick-gewichteten Websites im betrachteten Analysedatensatz, ist eine gute Klassifizierung des Geschlechts möglich. Es würden ungefähr

¹¹ Vgl. HTTP Archive (online verfügbar).

¹² Der verwendete Algorithmus XGBoost wird auch in der Unternehmenspraxis für vergleichbare Problemstellungen angewandt. Vgl. zum Beispiel die Website Kaggle (online verfügbar).

Kasten 2

Maschinelles Lernen und das Messen unternehmensspezifischer Skalenerträge

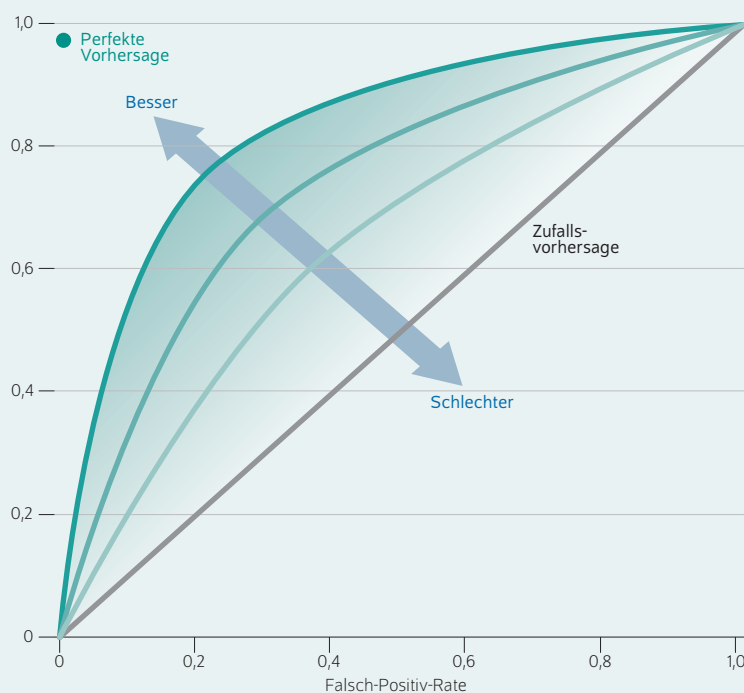
Maschinelles Lernen erlaubt Vorhersagen basierend auf Korrelationen eines vorherzusagenden Ergebnisses. Das kann beispielsweise die Zugehörigkeit einer Person zu einer bestimmten demografischen Gruppe sein. Für die Vorhersagen werden beobachtbare Informationen genutzt – im Falle dieser Studie sind das Internetnutzungsverläufe. Das Erstellen von Profilen für das Bereitstellen personalisierter Werbung entspricht sogenannten Klassifizierungsverfahren, in denen versucht wird, Internetnutzer*innen mittels maschinellen Lernens in verschiedene werberelevante Kategorien (zum Beispiel „weiblich“ als Konsumentinnenprofil) einzuteilen. In der Studie wird *extreme gradient boosting*, XGBoost, angewandt, ein gängiger Algorithmus für strukturierte Datensätze, welcher auf einer Ansammlung verschiedener Klassifizierungsmodelle basiert.

Die *Area Under the ROC Curve* (AUC) ermöglicht es, die Vorhersagequalität verschiedener Klassifikatoren anhand einer Kennzahl zu bemessen. Auf der *Receiver Operating Characteristic* (ROC)-Kurve wird die Richtig-Positiv-Rate (der Anteil richtig positiv klassifizierter Fälle aus allen positiven Fällen), der Falsch-Positiv-Rate (der Anteil falsch positiv klassifizierter Fälle aus allen nichtpositiven Fällen) einer Klassifizierungstechnologie gegenübergestellt. Die ROC-Kurve stellt alle machbaren Abwägungen zwischen der Richtig-Positiv-Rate und der Falsch-Positiv-Rate der Technologie dar. Die AUC liegt zwischen 0,5 (reine Zufallsklassifikation) und 1 (bestmögliches Klassifizierungsverfahren).¹

Um firmenspezifische Skalenerträge sowohl über die Anzahl von Nutzer*innen als auch über die Anzahl an beobachtbaren Websites zu messen, erstellt die Studie hypothetische Teildatensätze. Für jedes der zehn Klassifikationsprobleme wird hierfür die Detailtiefe (die Anzahl der Spalten) sowie die Größe (die Anzahl der Zeilen) des einem jeweiligen Datenbroker zu Verfügung stehenden Datensatzes systematisch variiert. Letzterer entspricht stets einem tabellarischen Datensatz, in welchem jede Zeile einem*iner Internetnutzer*in und jede Spalte einer Website zugeordnet ist. Zur Erstellung der Teildatensätze werden zufällig ausgewählte Internetnutzer*innen und Websites aus dem Datensatz eines Datenbrokers entfernt. Dies geschieht basierend auf einem vorher definierten Raster. Beispielsweise werden für große Datenbroker,

¹ Shan Huang, Michael Allan Ribers, Hannes Ullrich (2021): Der gesellschaftliche Mehrwert verknüpfter Daten: Algorithmen als Entscheidungshilfen bei Antibiotikaverschreibungen. DIW Wochenbericht Nr. 13/14, 239–246 (online verfügbar).

Abbildung

ROC-Kurve als Maß für Vorhersagequalität
Korrekt-Positiv-Rate

Anmerkungen: Beispielhafte Illustration. Die Korrekt-Positiv-Rate (Falsch-Positiv-Rate) bezeichnet den Anteil der korrekt (fälschlicherweise) positiv klassifizierten Fälle. Je weiter links oben die ROC-Kurve verläuft, desto besser ist die Vorhersagequalität. Die Diagonale ist die ROC-Kurve einer reinen Zufallsklassifikation (Münzwurf).

Quelle: Eigene Darstellung.

© DIW Berlin 2023

welche über 15 000 Webseiten tracken, pro Klassifikationsaufgabe jeweils ein Zehntel vom Ganzen der Internetnutzer*innen und Websites entfernt.

Auf Basis dieser Teildatensätze wird eine Simulationsanalyse durchgeführt, bei der für jede Klassifikationsaufgabe der Algorithmus zur Vorhersage einer Eigenschaft der Internetnutzer*innen angewandt wird. Die dabei erzielten Vorhersagequalitäten, gemessen durch die AUC, werden gesammelt und gespeichert. Dies ermöglicht die anschließende Quantifizierung von firmenspezifischen Skalenerträgen anhand von Regressionsanalysen.

95 Prozent der bestmöglichen Klassifizierung anhand von allen Daten erreicht werden. Das bedeutet: Auch mit eher wenigen Daten können die verschiedenen Dienste relativ gut Eigenschaften der Nutzer*innen anhand deren Verhaltens vorhersagen. Um eine weitere deutliche Steigerung der Klassifizierungsqualität zu erreichen ist aber ein so hoher Anteil an getrackten Websites nötig, wie ihn nur Google erreicht (Abbildung 4).

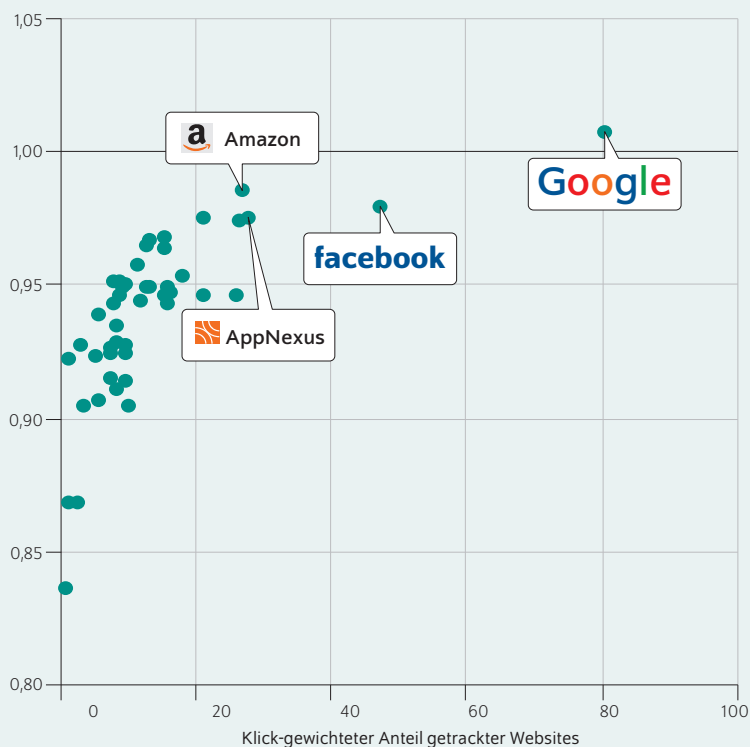
Google hat einen Wettbewerbsvorteil

Um Skalenerträge von Daten für jedes Unternehmen zu schätzen, wurde eine Simulationsanalyse durchgeführt, in der die Zahl der beobachteten Nutzer*innen und der beobachteten Websites variiert und mit der damit erreichbaren Klassifizierungsqualität verknüpft wird (Kasten 2). Hierauf basierend wird in einer Regressionsanalyse der

Abbildung 4

Vorhersagequalität der demografischen Profilbildung verschiedener Plattformen

Vorhersagequalität; Klick-gewichteter Anteil getrackter Websites in Prozent



Quelle: Eigene Berechnungen.

© DIW Berlin 2023

Die Datenmengen von Google führen zu einem Vorsprung bei der Vorhersagequalität der Nutzer*innenprofile.

Zusammenhang zwischen der Klassifizierungsqualität und der Anzahl der beobachteten Nutzer*innen sowie der beobachteten Websites geschätzt.

Sowohl für alle Firmen zusammengenommen als auch für die größten Firmen Google, Facebook, AppNexus und Amazon separat werden abnehmende Skalenerträge in der Anzahl an Nutzer*innen und Websites gemessen (Tabelle 1).

Die Ergebnisse für Google stechen jedoch heraus: Der Interaktionseffekt zwischen der Anzahl der Nutzer*innen und Websites hat einen positiven, statistisch signifikanten Koeffizienten. Das deutet darauf hin, dass Skalenerträge in der Anzahl an Nutzer*innen und Websites für Google weniger stark abnehmen als für andere Firmen. Somit können Daten allein einen nur schwer einholbaren Wettbewerbsvorteil für Google darstellen.

Fazit: Daten begünstigen eine konzentrierte Marktstruktur – Regulierung kann nötig sein

Marktführende Digitalunternehmen verfügen über enorme Datenmengen zu Verhalten und Vorlieben der

Verbraucher*innen, die sie zur Verbesserung ihrer Produkte und Dienstleistungen nutzen können. Durch den Zugang zu umfangreichen Datenbeständen können sie ihre Angebote zunehmend personalisieren, gezielte Werbung schalten und ihre Dienste kontinuierlich verbessern. Dies schafft einen Wettbewerbsvorteil gegenüber kleineren Unternehmen, die nicht über vergleichbare Ressourcen verfügen, um solche Daten zu sammeln und zu nutzen.

Die Untersuchung zeigt einen möglichen systematischen Unterschied in der Qualität der Bildung von Werbeprofilen von Unternehmen, die Nutzungsdaten sammeln. Die Ergebnisse suggerieren somit, dass das Teilen von Daten zwischen Online-Diensten ein wirksames Instrument sein kann, um Wettbewerb zu fördern und Innovationsanreize zu stärken. Der Digital Markets Act (DMA) der EU sieht vor, dass sogenannte Gatekeeper wie Alphabet und Meta, definiert auf Basis der Umsatzhöhe und Anzahl an Nutzer*innen, verpflichtet werden, Daten, die während der Nutzung ihrer Kerndienste entstehen, an Endnutzer*innen sowie gewerbliche Nutzer weiterzugeben.¹³ Der Durchsetzung dieser Regelung kommt daher für die Wettbewerbsdynamik insbesondere in der EU eine große Bedeutung zu.

Grundsätzlich wird die Profilbildung anhand von Third-Party-Cookies bereits durch die Datenschutzgrundverordnung der EU (EU-DSGVO) und zukünftig auch durch den DMA insbesondere in Europa zunehmend erschwert. Dennoch arbeiten die marktführenden Akteure der Online-Werbebranche an technologischen Entwicklungen, die Personalisierung und Kontextualisierung von Inhalten zwar datenschutzkonform ermöglichen, ihre Marktstellung bei den gegebenen Marktstrukturen aber weiter zementieren können. Dies betrifft nicht nur datenbezogene Strategien, sondern auch andere strategische Entscheidungen der marktbeherrschenden Plattformen, die zu einer Abschottung in Teilmärkten führen können.

Genau deswegen hat die Europäische Kommission zuletzt im Juni 2023 in einem bis dato einzigartigen Schritt angekündigt, auf die Marktstruktur abzielen und eine Zerschlagung der integrierten Werbedienste von Google zu erwirken, um den Wettbewerb von anderen Anbietern von Werbetechnologiediensten, Werbetreibenden und Online-Publishern zu ermöglichen.¹⁴

Diese Überlegungen sind nicht nur für den hier analysierten Markt von großer Relevanz, sondern im Allgemeinen für die aktuelle Wettbewerbsdynamik zwischen digitalen Plattformen in neu entstehenden Märkten. Zuletzt haben beispielsweise die dominierenden digitalen Unternehmen viele Ressourcen in die Entwicklung von großen Sprachmodellen

¹³ Gemäß dem DMA sind die Kerndienste von Online-Plattformen soziale Medien, Suchmaschinen, Videoplattformen, Kommunikationsdienste, intermediäre Dienstleistungen, Cloud-Dienste, Betriebssysteme und Werbe-Dienste. Auch der sich in der Abstimmung befindende europäische Data Act soll Unternehmen verpflichten, Daten mit anderen Unternehmen und Kund*innen zu teilen.

¹⁴ Pressemitteilung der Europäischen Kommission vom 14. Juni 2023: Antitrust: Commission sends Statement of Objections to Google over abusive practices in online advertising technology (online verfügbar).

Tabelle 1

Regressionsergebnisse der Simulationsanalyse

Variablen	(1) Alle Anbieter	(2) Google	(3) Facebook	(4) AppNexus	(5) Amazon
Anzahl der Nutzer*innen (in Tausend)	0,085***	0,084***	0,087***	0,041***	0,083***
Anzahl der Nutzer*innen (quadriert)	-0,012***	-0,012***	-0,011***	-0,004***	-0,011***
Anteil der getrackten Websites	0,420***	0,370***	0,306***	0,868***	0,430***
Anteil der getrackten Websites (quadriert)	-0,466***	-0,642***	-0,388**	-0,519	-1,426***
Anzahl der Nutzer*innen × Anzahl der Websites	0,008	0,038***	-0,010	-0,089**	0,036

Anmerkung: Die Sternchen an den Werten bezeichnen das statistische Signifikanzniveau. Je mehr Sternchen, desto geringer die Irrtumswahrscheinlichkeit: ***, ** und * geben die Signifikanz auf dem Ein-, Fünf- und Zehn-Prozent-Niveau an.

Lesehilfe: Die Vorhersagequalität ist relativ zur bestmöglichen Vorhersagequalität basierend auf dem vollständigen Datensatz definiert. Der Durchschnitt der so definierten Vorhersagequalität ist 0,86, was sich durch Teilen der durchschnittlichen absoluten Vorhersagequalität von 0,63 durch den Durchschnitt der bestmöglichen Vorhersagequalität von 0,73 ergibt. Die Zeilen 1 und 3 isoliert betrachtet ist für alle Anbieter (Spalte 1) ein Anstieg der Anzahl an Nutzer*innen um Tausend (Zeile 1) mit einem Anstieg der Vorhersagequalität um 8,5 Prozentpunkte assoziiert. Die Erhöhung des Anteils getrackter Websites (Zeile 3) um einen Prozentpunkt ist für Facebook (Spalte 3) mit einem Anstieg der Vorhersagequalität um 0,306 Prozentpunkte assoziiert. Aufgrund der quadratischen Terme mit negativen Werten in Zeilen 2 und 4 nehmen diese positiven Assoziationen allerdings zunehmend ab. Für Google (Spalte 2) wird diese Abnahme gedämpft, da ein gleichzeitiger Anstieg der Anzahl an Nutzer*innen um Tausend (Zeile 1) und des Anteils getrackter Websites (Zeile 3) positiv zur Vorhersagequalität beiträgt.

Quelle: Eigene Berechnungen.

© DIW Berlin 2023

investiert, die mittlerweile für die breite Masse zugänglich gemacht wurden. Für die Qualität dieser Modelle sind Daten und die Validierung durch Menschen essentiell. OpenAI hat bereits intern menschliche Bewertungen von generierten Inhalten verwendet, um das Modell GPT-4, das für den ChatGPT-Dienst verwendet wird, zu verbessern.¹⁵

Somit ist denkbar, dass das frühzeitige Bereitstellen von GPT-unterstützten Programmen durch Google, Meta und OpenAI/Microsoft ein frühzeitiges Verbessern der Algorithmen dank des Testens und Validierens durch Millionen von Nutzer*innen erlaubt. Wer diese Gelegenheit trotz potenzieller Risiken frühzeitig nutzt, kann so in Plattformmärkten Wettbewerbsvorteile generieren und langfristig Marktmacht zementieren. Ein verstärktes

Augenmerk auf diese Dynamik könnte die aktuellen Diskussionen um die Risiken von KI in konstruktivere und wirksamere Bahnen lenken als (wenn auch wichtige) Diskussionen über die, von manchen prophezeite, Verselbstständigung der KI.

Google und andere BigTech-Unternehmen konnten ihre Marktstellung und den hier gezeigten Wettbewerbsvorteil unter anderem gewinnen, weil diese Märkte über die letzten Jahrzehnte kaum ex-ante reguliert wurden und die ex-post wettbewerbsrechtliche Aufsicht nicht stark durchgesetzt wurde. Um in Märkten, für die die angesprochenen Sprachmodelle relevant sind, eine ähnliche Monopolisierung zu vermeiden, sollte am Beispiel der Entwicklung von Google gelernt werden. Frühe Eingriffe in diese Märkte können nötig werden, um Wettbewerbern den Zutritt zum Markt nicht zu versperren und weiterhin Innovationen zu ermöglichen.

¹⁵ Ajay Agrawal, Joshua Gans und Avi Goldfarb (2023): How Large Language Models Reflect Human Judgment. Harvard Business Review vom 12. Juni (online verfügbar).

Hannes Ullrich ist wissenschaftlicher Mitarbeiter der Abteilung Unternehmen und Märkte im DIW Berlin | hullrich@diw.de

Jonas Hannane ist Doktorand in der Abteilung Unternehmen und Märkte im DIW Berlin | jhannane@diw.de

Tomaso Duso ist Leiter der Abteilung Unternehmen und Märkte im DIW Berlin | tduso@diw.de

Christian Peukert war Fellow in der Abteilung Unternehmen und Märkte im DIW Berlin

Luis Aguiar ist Professor für Ökonomie an der Universität Zürich

JEL: D18, L40, L50, L86, M38

Keywords: Competition, Data Network Effects, Web Tracking, Platforms.

INTERVIEW



„Verfügbare Datenmengen können ein Unternehmen uneinholbar machen“

Hannes Ullrich, wissenschaftlicher Mitarbeiter der Abteilung Unternehmen und Märkte im DIW Berlin

1. **Herr Ullrich, in den letzten 25 Jahren ist ein Markt für personalisierte Werbung im Internet entstanden, der auf Tracking von Nutzer*innen und der Verwendung von Identifikatoren basiert. Sie haben sich mit diesem Markt auseinandergesetzt. Was stand dabei im Fokus Ihrer Untersuchung?** Wir wollten verstehen, welche Rolle die Verknüpfung von Daten für den Wettbewerb bedeuten kann, insbesondere, ob diese Verknüpfung von Daten aus verschiedenen Quellen zu den in der Vergangenheit beobachteten Monopolisierungstendenzen beigetragen haben könnte.
2. **Was sind die gängigsten Methoden, mit denen dabei gearbeitet wird?** Um Daten im Internet zu sammeln werden zum Beispiel bestimmte Codeschnipsel verwendet, die es zulassen, Personen über Websites hinweg zu verknüpfen. Dann werden, anhand dieser Daten Vorhersagen gebildet, zum Beispiel über die Präferenzen von Nutzer*innen. Das wird mit Hilfe statistischer Verfahren wie maschinellem Lernen gemacht.
3. **Wie hängen Qualität und Quantität der Daten zusammen? Bedeutet „mehr“ immer gleich „besser“?** „Mehr“ heißt schon im Grunde genommen immer „besser“, aber es heißt immer weniger „besser“. Wenn wir 1000 Datenpunkte haben und ich nehme 100 Datenpunkte dazu, dann ist das in der Regel besser. Wenn ich aber eine Million Datenpunkte habe und ich nehme 100 Datenpunkte dazu, dann ist das zwar ein bisschen besser, aber es ist im Endeffekt wahrscheinlich nicht mehr relevant für das Ziel, das ich verfolge. Diesen Zusammenhang nennt man „abnehmende Skalenerträge“. Es gibt bereits Studien, die dokumentieren, dass es diese abnehmenden Skalenerträge auch in der Qualität der Daten gibt. Aber was passiert, wenn ein Unternehmen sowohl die Menge als auch die Qualität der Daten erhöhen kann? Da hat uns interessiert, ob man in der Interaktion dieser beiden Arten von Skalen vielleicht Unterschiede finden kann, die dazu führen, dass bestimmte Unternehmen bessergestellt sind.
4. **Inwieweit haben die großen Player, wie Google, Amazon oder Facebook allein durch ihre schiere Größe Vorteile auf dem Datenmarkt?** Eine unserer Hypothesen war, dass es in diesem Markt sogenannte Datennetzwerkeffekte geben könnte. Datennetzwerkeffekte können entstehen, wenn ein Unternehmen Daten auf einer Website sammelt und dadurch den Nutzer*innen ein hochwertiges personalisiertes Produkt anbieten kann. Dafür sind dann Nutzer*innen bereit, mehr Zeit auf der Seite zu verbringen und vielleicht mehr Geld auszugeben. Dieses Geld geht an das Unternehmen, das diese Website veröffentlicht und ein Teil davon geht an das Unternehmen, das Daten sammelt. Da es jetzt mehr Geld gibt, erhöht sich vielleicht die inhaltliche Qualität dieser Website und Nutzer*innen sind mehr willens, diese Website zu besuchen. Dadurch können wiederum mehr Daten gesammelt werden, wodurch die Werbung wieder besser personalisiert wird. Damit schließt sich der Kreis und für andere Unternehmen wird es sehr schwer, dieses Unternehmen in der Qualität der Personalisierung einzuholen.
5. **Sollte der Datenmarkt reguliert werden und wenn ja, wie?** Es gibt ja schon einige Regulierungen, die auf dem Weg sind. In mindestens zweien davon ist vorgesehen, dass Unternehmen ihre Daten mit Wettbewerbern und mit Endnutzer*innen teilen sollen. Dieses Teilen von Daten kann es Unternehmen, die gar keine Chance haben, diese Daten zu sammeln, ermöglichen, doch in Konkurrenz zu treten, neue Produkte zu entwickeln und neue Märkte zu erschließen. Von daher kann eine Regulierung, die das Datenteilen erwirkt, durchaus zu mehr Wettbewerb, zu mehr Innovation und auch zu einer höheren Konsument*innenwohlfaht führen.

Das Gespräch führte Erich Wittenberg.



Das vollständige Interview zum Anhören finden Sie auf www.diw.de/interview

Discussion Papers Nr. 2038

2023 | Gert G. Wagner



Zur Geschichte des "arbeitnehmernahen" und "keynesianisch" geltenden Deutschen Instituts für Wirtschaftsforschung (DIW) seit den 50er Jahren

Für das im Jahr 2025 stattfindende 100-jährige Jubiläum des DIW Berlin und dessen Geschichte ist die hier vorgelegte Dokumentation eines 1999 geführten Gesprächs zwischen den ehemaligen DIW-Abteilungsleitern Fritz Franzmeyer und Reinhard Pohl einerseits und Gert G. Wagner andererseits von Interesse, das die Frage thematisiert, warum das Deutsche Institut für Wirtschaftsforschung (DIW Berlin) in der Öffentlichkeit seit

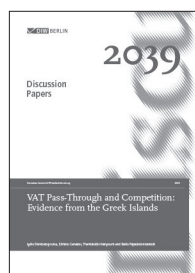
Jahrzehnten als "arbeitnehmernah" und "keynesianisch" etikettiert wurde. Die Dokumentation wurde bislang nicht veröffentlicht; die vorliegende Fassung wurde auf Basis eines eingescannten Ausdrucks im Mai 2023 fertiggestellt; inhaltliche Änderungen wurden nicht vorgenommen.

www.diw.de/publikationen/diskussionspapiere



Discussion Papers Nr. 2039

2023 | Lydia Dimitrakopoulou, Christos Genakos, Themistoklis Kampouris, Stella Papadokonstantaki



VAT Pass-Through and Competition: Evidence from the Greek Islands

We examine how competition affects VAT pass-through in isolated oligopolistic markets as defined by the Greek islands. Using daily gasoline prices and a difference-in-differences methodology, we investigate how changes in VAT rates are passed through to consumers in islands with different market structure. We show that pass-through increases with competition, going from 50% in monopoly to around 80% in more competitive markets, but remains incomplete. We also discover a rapid rate of adjustment for VAT changes, as well as a positive relationship between competition and the rate of price adjustment. Finally, we

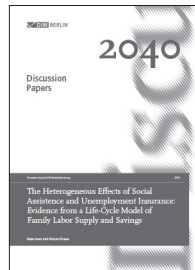
document higher pass-through for products with more inelastic demand.

www.diw.de/publikationen/diskussionspapiere



Discussion Papers Nr. 2040

2023 | Peter Haan, Victoria Prowse



The Heterogeneous Effects of Social Assistance and Unemployment Insurance: Evidence from a Life-Cycle Model of Family Labor Supply and Savings

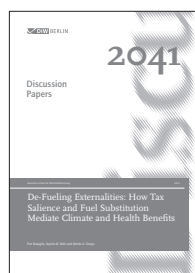
We empirically analyze the heterogeneous welfare effects of unemployment insurance and social assistance. We estimate a structural life-cycle model of singles' and married couples' labor supply and savings decisions. The model includes heterogeneity by age, education, wealth, sex and household composition. In aggregate, social assistance dominates unemployment insurance; however, the opposite holds true for married men, whose leisure time declines more than that of their spouses when unemployment insurance is reduced. A revenue-neutral rebalancing of social support away from unemployment insurance and toward social assistance increases aggregate welfare. Income pooling in married households decreases the welfare value of social assistance.

www.diw.de/publikationen/diskussionspapiere



Discussion Papers Nr. 2041

2023 | Pier Basaglia, Sophie Behr, Moritz A. Drupp



De-Fueling Externalities: How Tax Salience and Fuel Substitution Mediate Climate and Health Benefits

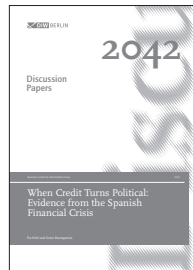
This paper is the first to investigate the effectiveness of fuel taxation to jointly deliver climate and health benefits in a quasi-experimental setting. Using the synthetic control method, we compare carbon and air pollutant emissions of the actual and synthetic German transport sector following the 1999-2003 German eco tax reform. We demonstrate sizable average reductions in CO₂ (12%), PM_{2.5} (10%) and NO_x (6%) emissions between 1999 and 2009 across a range of specifications. Using official cost estimates, we find that the eco-tax saved more than 40 billion euros of external damages. More than half of the reductions in external damages are health benefits, highlighting the importance of accounting for co-pollution impacts of carbon pricing. Our fuel and emission specific tax elasticity estimates suggest much stronger demand responses to eco tax increases than to market price movements, primarily due to increases in tax salience, which we measure using textual analysis of newspapers. We further show that gasoline-to-diesel substitution substantially mediates the trade-off between climate and health benefits. Our results highlight the key roles of tax salience and fuel-substitution in mediating the effectiveness of fuel taxes to reduce climate and health externalities.

www.diw.de/publikationen/diskussionspapiere



Discussion Papers Nr. 2042

2023 | Pia Hüttl, Simon Baumgartner



When Credit Turns Political: Evidence from the Spanish Financial Crisis

This paper provides causal evidence on the effect of credit crunches on political polarization. We combine data on bank-firm connections and electoral outcomes at the city-level during the 2008-2014 Spanish Financial Crisis. First, we show that firms in a relationship with weak banks experience a reduction in their loan supply and employment growth. Next, we estimate the effects of unemployment on voting behaviour. We construct an instrument for unemployment based on the city-level exposure to foreign weak banks. We find that a one standard deviation increase in instrumented unemployment translates into

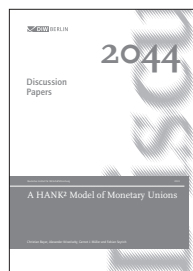
a 7 percentage increase in the polarisation of voters.

www.diw.de/publikationen/diskussionspapiere



Discussion Papers Nr. 2044

2023 | Christian Bayer, Alexander Kriwoluzky, Gernot J. Müller, Fabian Seyrich



A HANK2 Model of Monetary Unions

How does a monetary union alter the impact of business cycle shocks at the household level? We develop a Heterogeneous Agent New Keynesian model of two countries (HANK2) and show in closed form that a monetary union shifts the adjustment to a shock horizontally—across countries—within the brackets of the union-wide wealth distribution rather than vertically—that is, across the brackets of the union-wide wealth distribution. Calibrating the model to the euro area reveals that a monetary union alters the impact of shocks most strongly in the tails of the wealth distribution but leaves the middle class almost unaffected.

www.diw.de/publikationen/diskussionspapiere



CLAUDIA KEMFERT UND CHRISTIAN VON HIRSCHHAUSEN



Claudia Kemfert, Leiterin
der Abteilung Energie,
Verkehr, Umwelt im
DIW Berlin.



Christian von Hirschhausen,
Forschungsdirektor
Abteilung Energie, Verkehr,
Umwelt im DIW Berlin.

Bundesregierung überschätzt Bedarf an Flüssiggas- Infrastruktur

Der Kommentar gibt die Meinung der Autor*innen wieder.

Trotz heftiger Bürger*innenproteste unter anderem auf der Insel Rügen hält die Bundesregierung an ihrem Vorhaben fest, die Flüssiggas-Infrastruktur an der Ost- und Nordsee auszubauen und sogar zu beschleunigen. Das Gesetz könnte unmittelbar vor der Sommerpause im Bundestag beschlossen werden. Über den Sinn des Gesetzes lässt sich streiten.

Tatsächlich hat die Bundesregierung, allen voran Wirtschaftsminister Robert Habeck, im vergangenen Jahr durch den Ankauf schwimmender Flüssiggasterminals zu der Beruhigung der Lage an den Gasmärkten beigetragen. Dem gilt Respekt. Heute ist jedoch absehbar, dass der weitere Ausbau dieser Infrastruktur nicht nur energiewirtschaftlich nicht notwendig, sondern auch mit erheblichen Kosten verbunden ist, die in der aktuellen Diskussion weitgehend unter den Tisch gekehrt werden. Tatsächlich hat sich die Lage entspannt und die Argumente, die vor einem Jahr noch Gültigkeit gehabt haben mögen, sind nicht mehr einschlägig.

So ist die befürchtete Gasmangellage ausgeblieben. Die Gaspreise haben sich stabilisiert, sowohl in Deutschland als auch in Europa. Sowohl Industrie- als auch Haushalte haben sich als sehr preiselastisch erwiesen, das heißt, sie haben erhebliche Effizienzpotenziale genutzt und den Verbrauch im Jahr 2022 um circa 14 Prozent gesenkt. Dies sollte in Zukunft auch weitergeführt werden. Fabrikausfälle und kalte Nächte sind ausgeblieben. Und auch der kommende Winter 2023/24 wird entspannt bleiben: Angesichts auskömmlicher Gaslieferungen und prall gefüllter Speicher ist keine Gasmangellage absehbar.

Der im LNG-Beschleunigungsgesetz vorgesehene Bau fester Infrastruktur ist weder energiepolitisch notwendig noch klimapolitisch sinnvoll. Zu diesem Ergebnis kommen die meisten Studien aus unterschiedlichen Richtungen, wie dem Energiewirtschaftlichen Institut der Universität zu Köln (ewi), dem New Climate Institute und dem DIW Berlin. In den ersten Kriegsmontaten nach dem 24. Februar 2022 kam es zu einer Dynamik, die die Gaswirtschaft zur Beantragung von Projekten weit jenseits sinnvoller Mengen genutzt hat. Die Rede ist von 30 Mil-

liarden Kubikmetern. Dagegen zeigen Modellrechnungen, dass bereits Ende der 2020er Jahre der Einsatz von Flüssiggas weitgehend überflüssig sein wird. Weiterer Infrastrukturausbau muss somit als „stranded assets“, gestrandete Investitionen, gelten. Steuerzahlende und Wirtschaft werden dagegen durch die abgeschlossenen Lieferverträge noch langfristig belastet.

Die Transitfunktion Deutschlands, inklusive in Richtung Mittel- und Südeuropa, ist gesichert. Insbesondere bedarf es für die Versorgung der ostdeutschen Bundesländer keiner zusätzlichen Importterminals an der Ostsee. Das deutsche Gasnetz ist gut ausgebaut und bei rückläufiger Nachfrage sind auch in Zukunft keine Transportengpässe zu befürchten. Die europäische Solidarität wurde 2022 gelebt und ist auch mit der bestehenden Infrastruktur für die kommenden Jahre darstellbar.

Aus der Analyse lassen sich drei Schlüsse ziehen: Die Erdgasmärkte sollten aus der staatlichen Bewirtschaftung der (nicht eingetretenen) Gasmangellage befreit werden. Der beschleunigte Ausbau von LNG-Importinfrastruktur an Ost- und Nordsee sollte zeitnah abmoderiert werden. Insbesondere ist dieser Aufbau in Mukran auf Rügen, der im Zentrum der Diskussion um das LNG-Beschleunigungsgesetz steht, nicht für die Versorgungssicherheit der neuen Bundesländer oder der EU-Nachbarländer notwendig. Last but not least sollten die Haushalte*innen im Deutschen Bundestag kritischer mit den bisher reichlich sprudelnden Mitteln sein: Noch besteht Gelegenheit, die Beschleunigung des teuren und unnötigen LNG-Infrastrukturausbaus zu stoppen und die Energiewende in Richtung Erdgasausstieg weiterzutreiben. Der ökonomisch und ökologisch sinnvolle Erdgasausstieg sollte konsequent weitergeführt werden, statt ihn durch zusätzliche fossile Infrastruktur zu verzögern.

Dieser Kommentar ist in einer längeren Version im Tagesspiegel Background „Energie & Klima“ erschienen.