Lecture 11:
(i) Biased algorithms;
(ii) Privacy

CMSC 25900 / DATA 25900

Spring 2021

The University of Chicago

THE UNIVERSITY OF CHICAGO

Biased algorithms (and biased data)

Unsupervised Models Are Biased, Too!

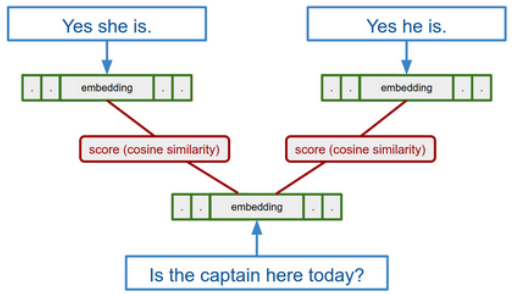
- https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html?m=1

As Machine Learning practitioners, when faced with a task, we usually select or train a model primarily based on how well it performs on that task. For example, say we're building a system to classify whether a movie review is positive or negative. We take 5 different models and see how well each performs this task:

Figure 1: Model performances on a task. Which model would you choose?

Normally, we'd simply choose Model C. But what if we found that while Model C performs the best overall, it's also most likely to assign a more positive sentiment to the sentence "The main character is a man" than to the sentence "The main character is a woman"? Would we reconsider?

Gender Biases of Chatbots



	score
Yes she is	0.17
Yes he is	0.27
Bias score (difference)	-0.10

For a given occupation overall, the model's bias score is the sum of the bias scores for all question/answer templates with that occupation.

Tamera runs 200 occupations through this analysis using the Universal Sentence Encoder embedding model. Table 2 shows the occupations with the highest female-biased scores (left) and the highest male-biased scores (right):

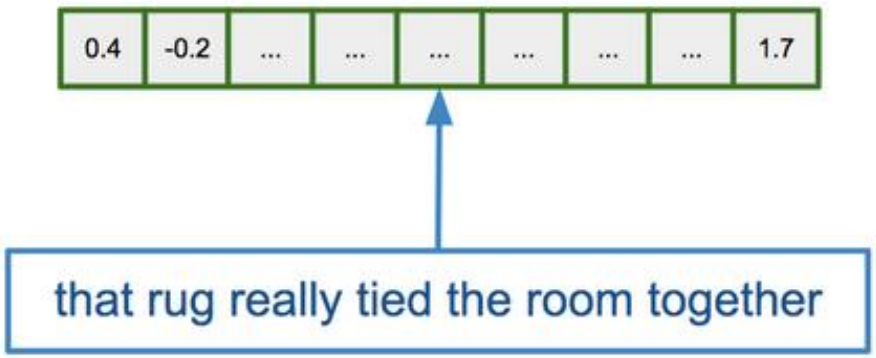
Highest female bias				Highest male bias			
occupation	bias	occupation	bias	occupation	bias	occupation	bias
maid	59.2	librarian	20.1	undertaker	-73.4	captain	-53.4
waitress	52.5	obstetrician	16.9	janitor	-62.3	announcer	-51.1
midwife	50.9	secretary	13.7	referee	-60.7	architect	-50.7
receptionist	50.2	socialite	12.1	plumber	-58	maestro	-50.6
nanny	47.7	therapist	10.2	actor	-56.9	drafter	-46.7
nurse	45.4	manicurist	10.1	philosopher	-56.2	usher	-46.6
midwives	43.8	hairdresser	9.7	barber	-55.4	farmer	-45.4
housekeeper	36.6	stylist	8.6	umpire	-54.3	broadcaster	-45.2
hostess	32	homemaker	6.9	president	-54	engineer	-45.1
gynecologist	31.6	planner	5.8	coach	-53.8	magician	-44.8

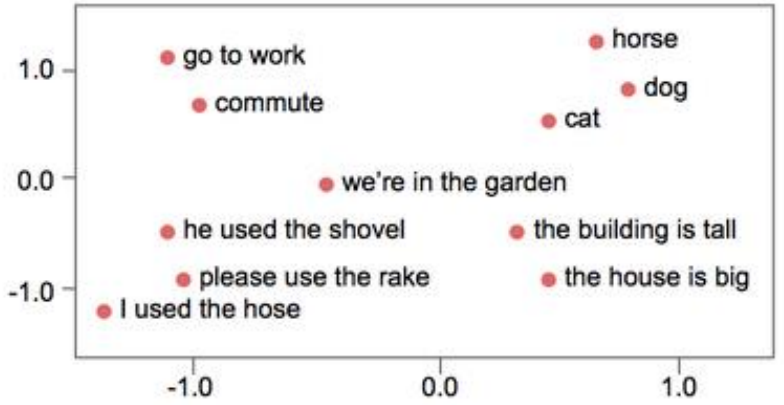
Table 2: Occupations with the highest female-biased scores (left) and the highest male-biased scores (right).

https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html?m=1

Word Embeddings

Text embedding models convert any input text into an output vector of numbers, and in the process map semantically similar words near each other in the embedding space:





https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html?m=1

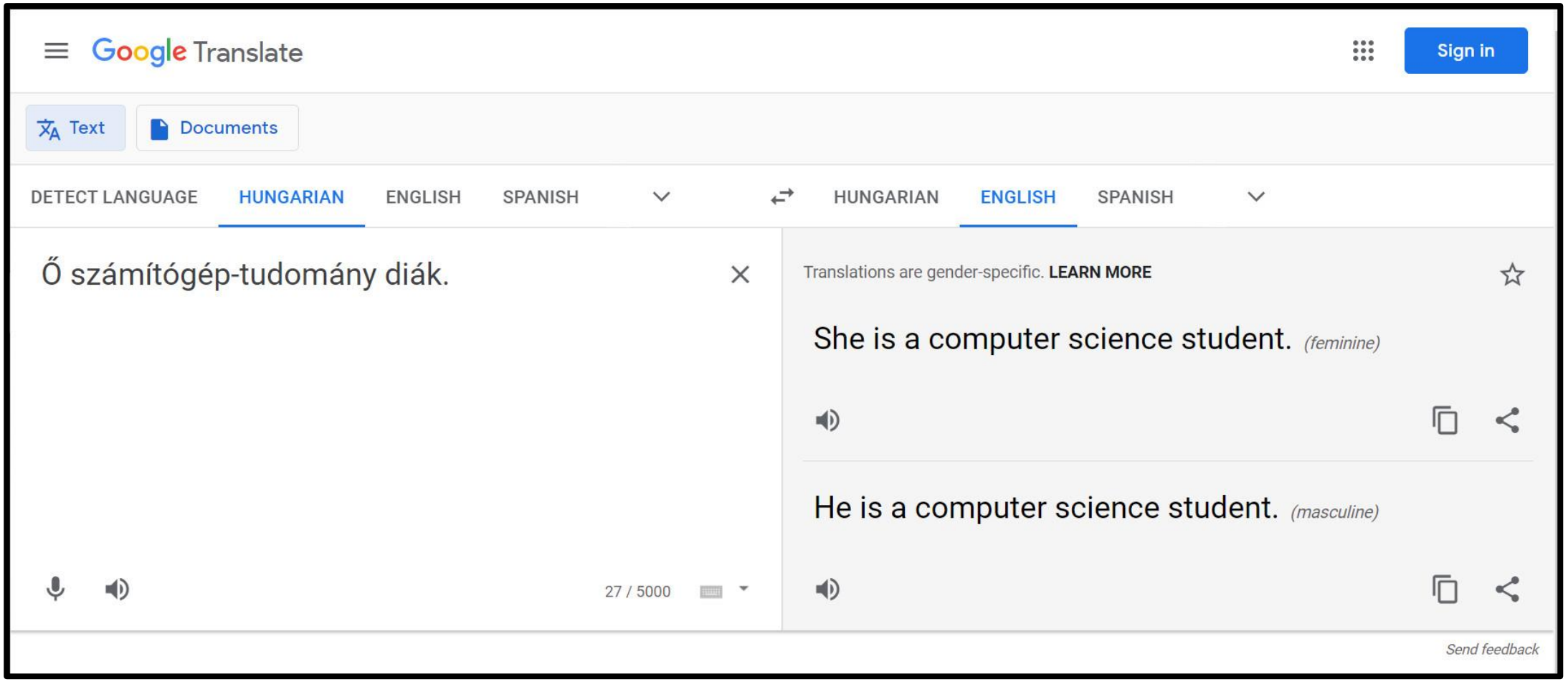
Gender Biases of Chatbots

Targets (N)	Attributes (N)	GloVe*	word2vec	nnlm-en-dim50	nnlm-en-dim128	universal
Flowers vs Insects (25)	Pleasant vs Unpleasant (25)	**1.50***	**1.54***	**1.54***	**1.63***	**1.38***
Instruments vs Weapons (25)	Pleasant vs Unpleasant (25)	**1.53***	**1.63***	**1.66***	**1.55***	**1.44***
Eur-American vs Afr-American Names[6] (25)	Pleasant vs Unpleasant[6] (25)	**1.41***	**0.58***	**0.70***	0.04	0.36
Eur-American vs Afr-American Names[7] (18)	Pleasant vs Unpleasant[6] (25)	**1.50***	**1.24***	**1.04***	0.23	-0.37
Eur-American vs Afr-American Names[7] (18)	Pleasant vs Unpleasant[8] (8)	**1.28***	**0.72***	0.28	-0.09	0.72
Male vs Female names (8)	Career vs Family (8)	**1.81***	**1.89***	**1.45***	**1.70***	0.03
Math vs Arts (8)	Male vs Female (8)	1.06	0.97	**1.29***	1.07	0.59
Mental vs Physical Disease (6)	Temporary vs Permanent (7)	**1.38***	1.30	**1.35***	0.96	**1.60***
Science Arts (8)	Male vs Female (8)	**1.24***	**1.24***	**1.34***	1.19	0.24
Young vs Old Names (8)	Pleasant vs Unpleasant (8)	**1.21**	-0.08	0.75	-0.47	1.01

*Table 1: Word Embedding Association Test (WEAT) scores for different embedding models. Cell color indicates whether the direction of the measured bias is in line with (blue) or against (yellow) the common human biases recorded by the Implicit Association Tests. *Statistically significant ($p < 0.01$) using Caliskan et al. (2015) permutation test. Rows 3-5 are variations whose word lists come from [6], [7], and [8]. See Caliskan et al. for all word lists. * For GloVe, we follow Caliskan et al. and drop uncommon words from the word lists. All other analyses use the full word lists.*

https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html?m=1

Gender in Language



Data Encapsulates Values

https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation

Representation Matters in Data

Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Table 4: Gender classification performance as measured by the positive predictive value (PPV), error rate (1-PPV), true positive rate (TPR), and false positive rate (FPR) of the 3 evaluated commercial classifiers on the PPB dataset. All classifiers have the highest error rates for darker-skinned females (ranging from 20.8% for Microsoft to 34.7% for IBM).

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proc. FAT**, 2018. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Representation Matters in Data

Figure 3: The percentage of darker female, lighter female, darker male, and lighter male subjects in PPB, IJB-A and Adience. Only 4.4% of subjects in Adience are darker-skinned and female in comparison to 21.3% in PPB.

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proc. FAT**, 2018. http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Privacy

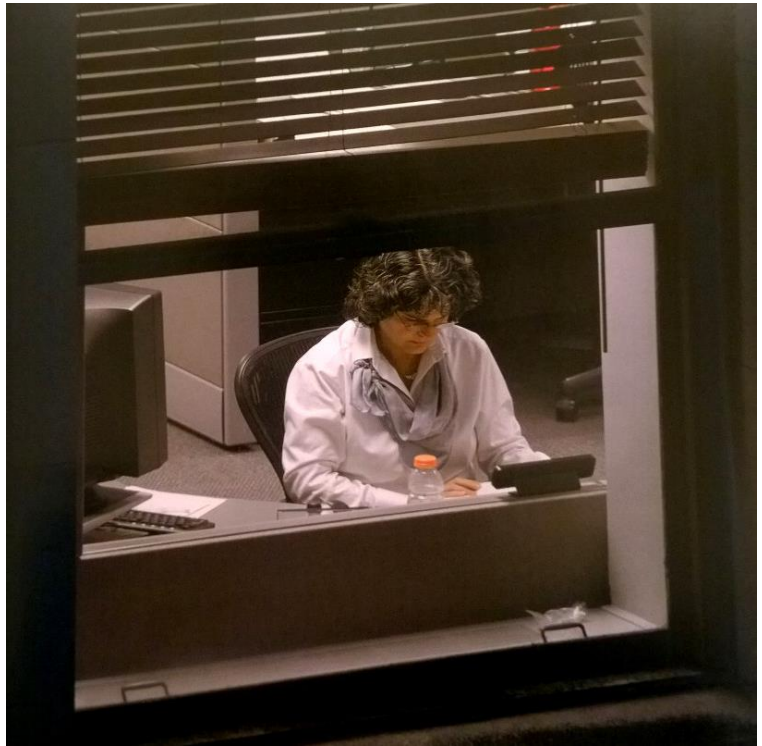
Privacy is Hard to Define

“Privacy is a value so complex, so entangled in competing and contradictory dimensions, so engorged with various and distinct meanings, that I sometimes despair whether it can be usefully addressed at all.”

Robert C. Post, Three Concepts of Privacy,
89 Geo. L.J. 2087 (2001).

Michael Wolf- The Transparent City

Michael Wolf- The Transparent City



"Chicago has recently undergone a surge of new construction…In early 2007, the Museum of Contemporary Photography…invited Michael Wolf as an artist-in-residence….Wolf chose to photograph the central downtown area, focusing on issues of voyeurism and the contemporary urban landscape….his details are fragments of life—digitally distorted and hyper-enlarged—snatched surreptitiously via telephoto lenses

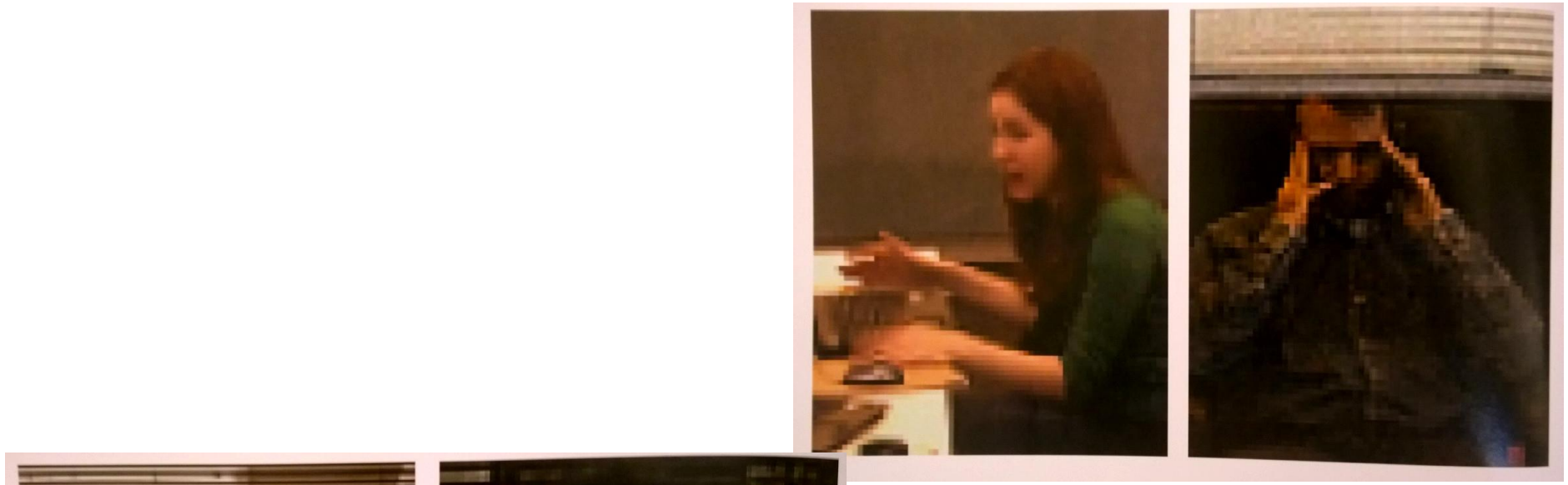
http://aperture.org/shop/the-transparent-city/

Michael Wolf- The Transparent City

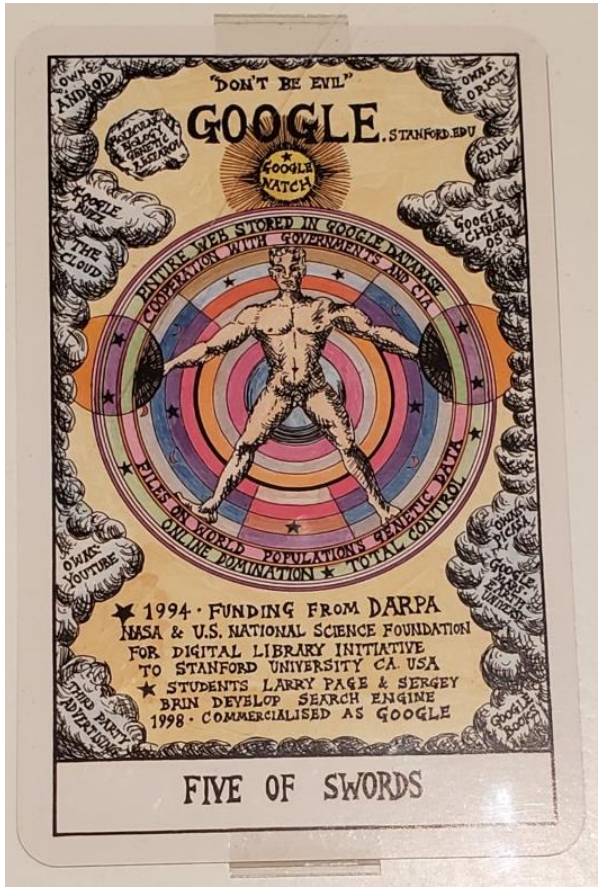
Michael Wolf- The Transparent City

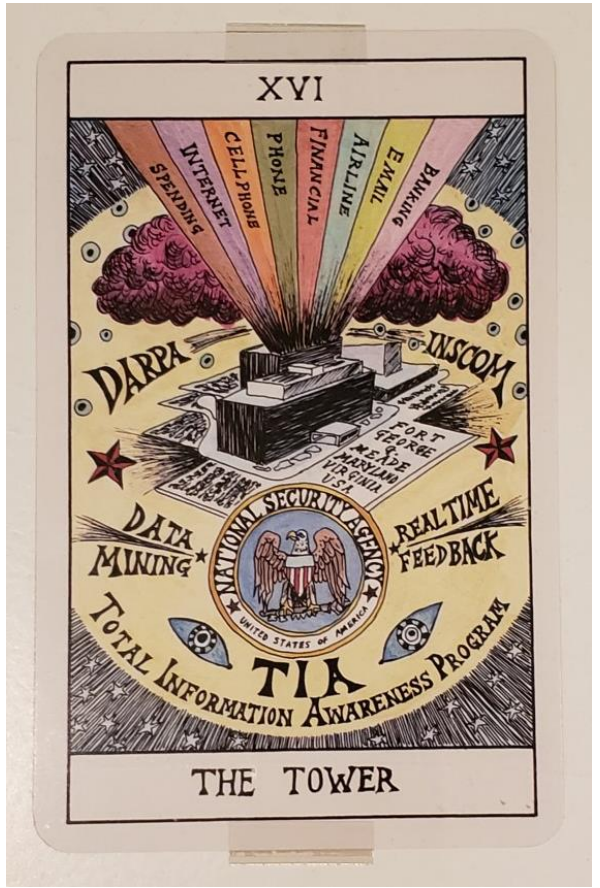
Michael Wolf- The Transparent City

Michael Wolf- The Transparent City

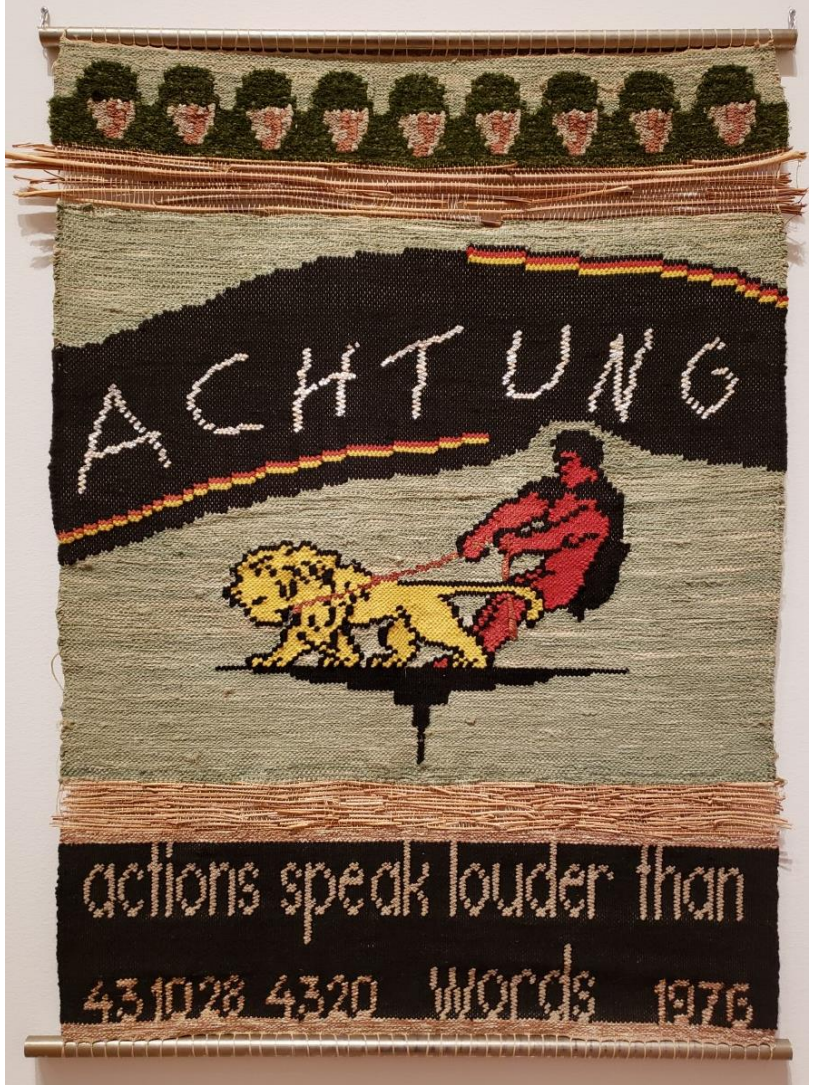


Suzanne Treister: HEXEN 2.0 Tarot (2011)





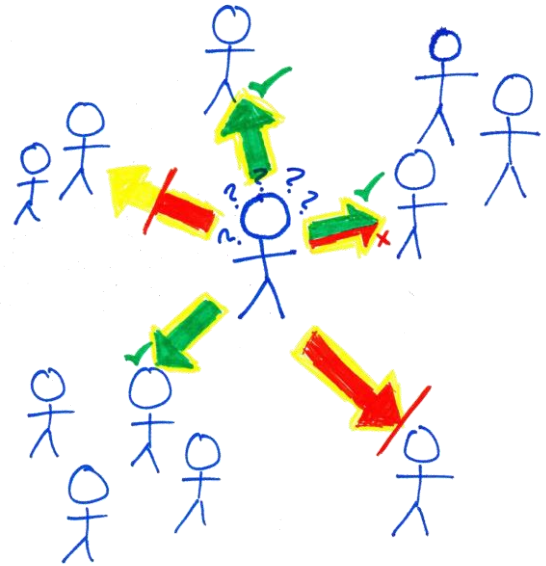
Charlotte Johanesson: Achtung (1976)



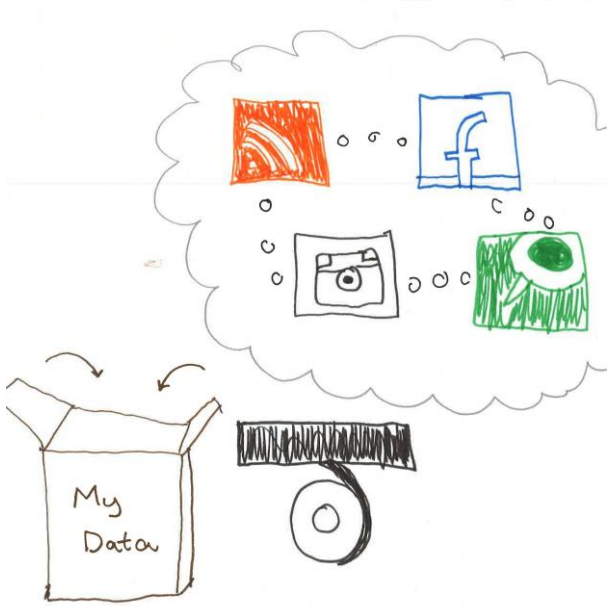
Nam June Paik

Rafael Lozano-Hemmer: Zoom Pavilion (2015)

Rafael Lozano-Hemmer: Zoom Pavilion (2015)



Conceptualizing Privacy



Warren and Brandeis (1890)

HARVARD

LAW REVIEW.

VOL. IV. DECEMBER 15, 1890. NO. 5.

THE RIGHT TO PRIVACY.

"It could be done only on principles of private justice, moral fitness, and public convenience, which, when applied to a new subject, make common law without a precedent; much more when received and approved by usage."

WILLES, J., in Millar *v.* Taylor, 4 Burr. 2303, 2312.

THAT the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection. Political, social, and economic changes entail the recognition of new rights, and the common law, in its eternal youth, grows to meet the demands of society. Thus, in very early times, the law gave a remedy only

Warren and Brandeis's Inspiration

Warren and Brandeis's Argument

- "The individual shall have full protection in person and in property"
- The legal basis for fear
  - Battery → assault
  - Tangible property → intangible property
- Gossip pages about high society

Warren and Brandeis's Argument

- Libel and slander are insufficient in considering only damage to reputation
- Considers property rights
- The right to prevent, rather than profit from, publication
- **"The right to be let alone"**
- Excludes topics of general interest

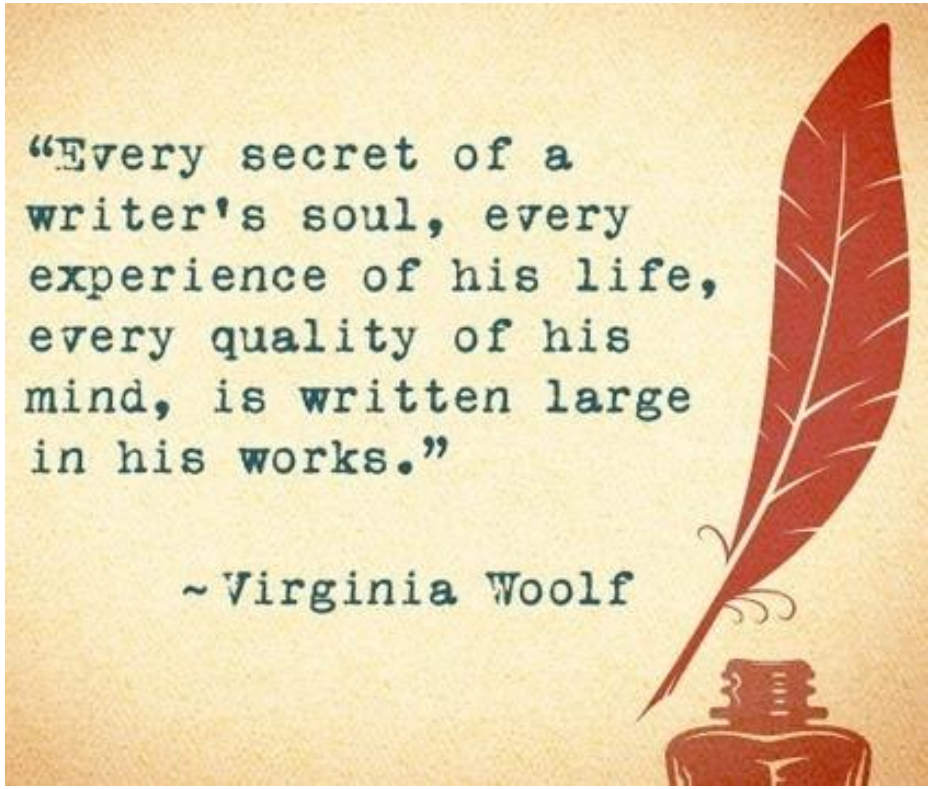
Photography Laws

Consent required for action related to a picture of a person in a public place (by country)

Country	Take a picture	Publish a picture	Commercially[1] use a published picture
Afghanistan	No	Yes (with exceptions)	Yes (with exceptions)
Argentina	No	Yes (with exceptions)	Yes (with exceptions)
Australia	No (with exceptions)	No (with exceptions)	Yes
Austria	No	No (with exceptions)	Yes
Belgium	No	Yes (with exceptions)	Yes
Brazil	Yes	Yes	Yes
Bulgaria	No	No	Yes
Canada	Depends on province	Yes (with exceptions)	Yes
China	No	No	Yes
Czech Republic	Yes (with exceptions)	Yes (with exceptions)	Yes (with exceptions)
Denmark	No	Yes (with exceptions)	Yes (with exceptions)
Ethiopia	No	Yes (with exceptions)	Yes
Finland	No	Yes (with exceptions)	Yes (with exceptions)
France	Yes (with exceptions)	Yes (with exceptions)[3]	Yes
Germany	No (with exceptions)	Yes (with exceptions)	Yes (with exceptions)
Greece	No	No	Yes (with exceptions)
Hong Kong	Depends on circumstances	Depends on circumstances	Depends on circumstances
Hungary	Yes (with exceptions)	Yes (with exceptions)	Yes (with exceptions)
United Kingdom	Depends on circumstances	Depends on circumstances	Depends on circumstances
United States	No	No	Usually (although laws differ by state)

https://commons.wikimedia.org/wiki/Commons:Photographs_of_identifiable_people#The_right_of_publicity

Is Being “Let Alone” Sufficient?

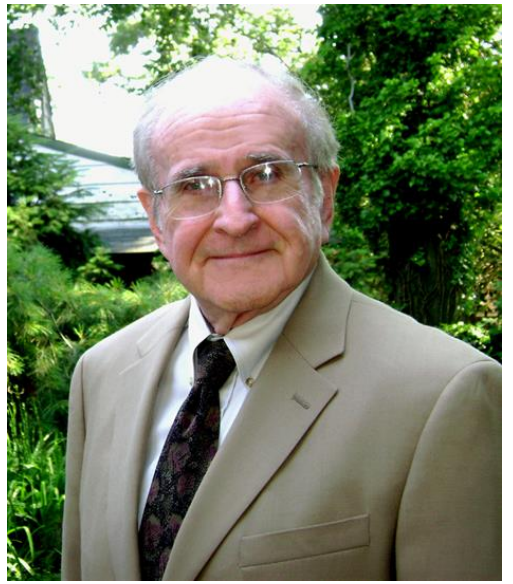


Privacy as Control / Secrecy (1967)

“Privacy is the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.”

“…each individual is continually engaged in a personal adjustment process in which he balances the desire for privacy with the desire for disclosure and communication….”

Alan Westin, Privacy and Freedom, 1967



Is Limiting Access Sufficient?

- Individuals sometimes prefer to be let alone, yet sometimes want to be social
  - Privacy was traditionally “social withdrawal”

Privacy Regulation Theory (1975)

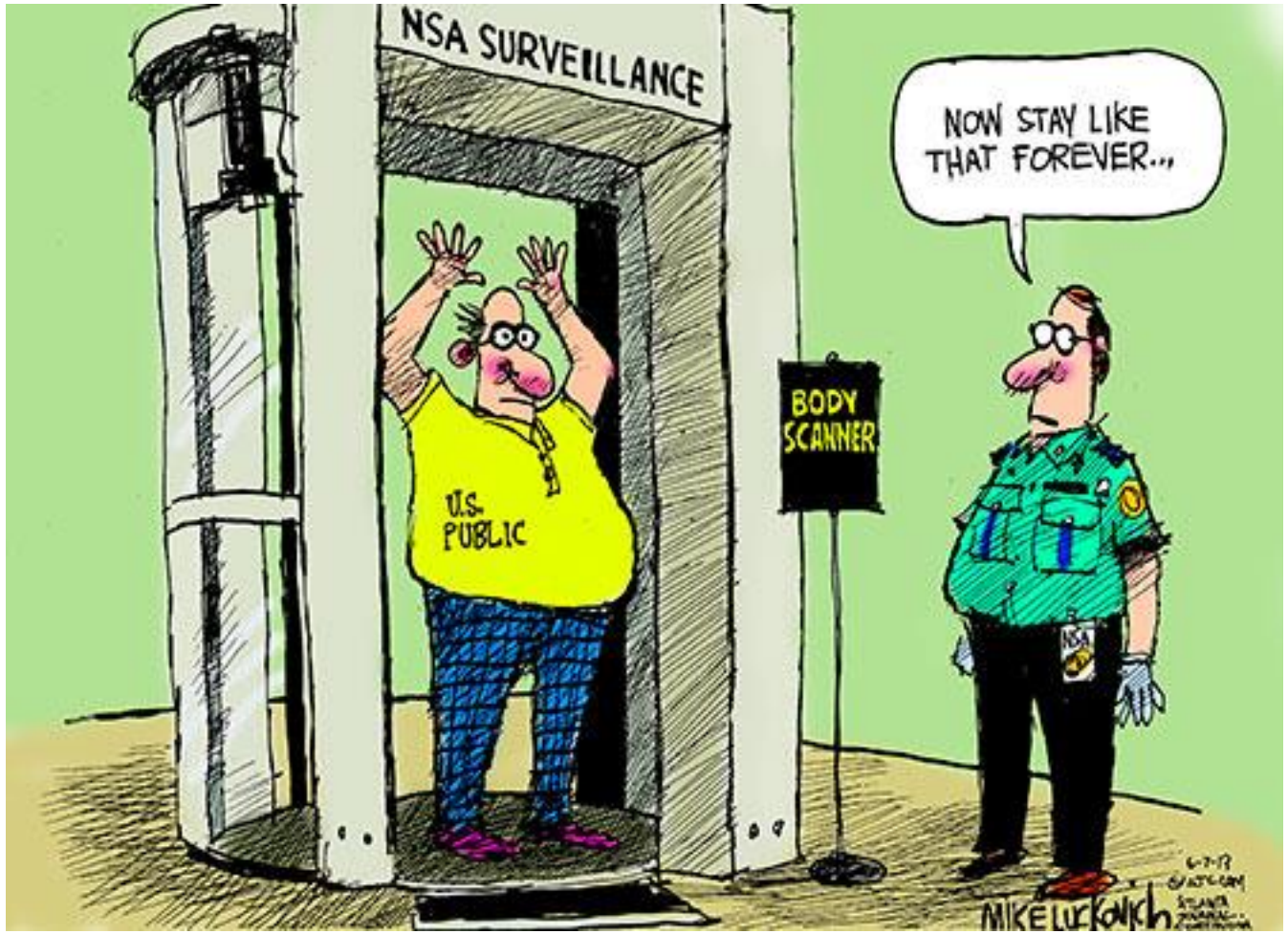
- Irwin Altman (social psychology)
  - Preceded by Altman and Taylor's Social Penetration Theory (1973) about intimacy in relationships
- Dialectic and dynamic process of boundary regulation
  - Continuous movement on a continuum
- Goal: optimum balance of privacy and social interaction

CPM Theory (1991)

- Sandra Petronio (communications)
  - Communication Privacy Management Theory
- Regulate boundaries based on perceived costs and benefits
  - Movement on a continuum
- Expect rule-based management
- Boundary turbulence related to clashing expectations

Is Regulating Disclosure Enough?

Purpose Matters



Privacy as Contextual Integrity (2004)

- Helen Nissenbaum (philosophy)
- “Contextual integrity ties adequate protection for privacy to norms of specific contexts, demanding that information gathering and dissemination be appropriate to that context.”

Privacy as Contextual Integrity (2004)

- Appropriate flows of information
- Appropriate flows conform to contextual information norms
- Norms refer to the data subject, sender, recipient, information type, and transmission principle
- Conceptions of privacy evolve over time and are grounded in ethics

Dan Solove's Pluralistic Conceptions

- Some data isn't "sensitive," but its collection and use impact privacy
  - Impact power relationships
  - Kafka-esque
- Solove's privacy taxonomy
  - Information collection
  - Information processing
  - Information dissemination
  - Invasion

Privacy in Practice

Technology

Apple employee detained by U.S. customs agents after declining to unlock phone, laptop

Customs and Border Protection officers violated a citizen's rights when they demanded he turn over passwords to his electronic devices at the airport, the American Civil Liberties Union Foundation of Northern California said in a civil complaint filed Tuesday. (Daniel Acker/Bloomberg)

By **Hamza Shaban**
April 3

When Andreas Gal returned from a business trip in Sweden last fall, he was carrying two company-owned devices: an iPhone XS that flashed "Confidential and Proprietary" on its lock screen and a MacBook Pro bearing a sticker that read "PROPERTY OF APPLE. PROPRIETARY."

CHICAGO IS TRACKING KIDS WITH GPS MONITORS THAT CAN CALL AND RECORD THEM WITHOUT CONSENT

Cook County has a new contract for juvenile ankle monitors that critics say are an invasion of privacy.

This story was co-published with Citylab.

On March 29, court officials in Chicago strapped an ankle monitor onto Shawn, a 15-year-old awaiting trial on charges of armed robbery. They explained that the device would need to be charged for two hours a day and that it would track his movements using GPS technology. He was told he would have to be given permission to leave his house, even to go to school.

Privacy-related terminology

- **Chilling effect**: discouragement of exercising a legitimate right
- **Privacy paradox**: behaviors appear inconsistent with concerns
- **Privacy by design**: consider privacy throughout the lifecycle of a product
- **Secondary use**: those other than the intended purpose

Issues of privacy

- Can conflict with free speech / security
- How do we quantify privacy harms?
- Can we measure chilling effects?
- How do we provide transparency?
- Distortion: false or misleading information
- Data mining → future activities?
- Oversight and accountability

Surveillance

Measuring privacy

- Why is privacy hard to measure?
- Why are attitudes about privacy hard to measure?
- Why is the cost of privacy invasion hard to measure?

Privacy Law and Regulation

How privacy is protected

- Laws, self regulation, technology
  - Notice and access
  - Control over collection, use, deletion, sharing
  - Collection limitation
  - Use limitation
  - Security and accountability

OECD Fair Information Principles

- Collection limitation
- Data quality
- Purpose specification
- Use limitation
- Security safeguards
- Openness
- Individual participation
- Accountability
- http://www.privacyrights.org/ar/fairinfo.htm

US FTC’s Fair Information Practice Principles (FIPPs)

- Notice / Awareness
- Choice / Consent
- Access / Participation
- Integrity / Security
- Enforcement / Redress
- https://en.wikipedia.org/wiki/FTC_Fair_Information_Practice

Right to be forgotten

- Should a person have the agency to cause items from the past to be removed?
- Who owns information?

Privacy on the books / on the ground

- Data Protection Directive (1995, since superseded by GDPR) - EU countries must adopt similar comprehensive laws, recognize privacy as fundamental human right
  - Privacy commissions in each country
- US has sector-specific laws, minimal protections, “patchwork quilt”
  - No explicit constitutional right to privacy or general privacy law
  - Some privacy rights inferred from constitution
  - Narrow regulations for health (HIPAA 1996), credit (FCRA 1970), education (FERPA 1974), video rental records (VPPA 1998), children (COPPA 1998)
  - FTC investigates **unfair** & **deceptive** practices
  - FCC regulates telecommunications
  - Some state and local laws

General Data Protection Regulation (2016)

- **GDPR** came into effect May 25, 2018 and applies to the EU
- Distinguishes between data subjects, controllers (people who direct analysis), and processors (those who do the analysis)
- Data controller informs the 'data subject in a concise, **transparent**, intelligible and easily accessible form, using **clear and plain language**'
- **Right of access** for data subjects
- **Right of erasure** (with some exceptions)
- **Right to object** to processing for some purposes
- **Privacy by design** (Article 25)

General Data Protection Regulation (2016)

- Pseudonymization required for stored personal data
- Data breach notification to authorities within 72 hours
- Possible fines of up to 4% of worldwide turnover
- Can only process data based on six lawful bases:
  - Consent
  - Contract
  - Public task
  - Vital interest
  - Legitimate interest
  - Legal requirement

California Consumer Privacy Act (2018)

- **CCPA** went into effect January 1, 2020
- Residents of California have rights to:
  - Know what personal data is collected
  - Know whether that data is sold
  - Refuse the sale of personal data
  - Access their data
  - Request erasure of their personal data
  - Not be discriminated against for exercising these privacy rights
- Fine of $7,500 for intentional and $2,500 for unintentional violations

Virginia Consumer Data Protection Act (2021)

iapp

Virginia passes the Consumer Data Protection Act

Mar 3, 2021 Save This

Sarah Rippy
IAPP Staff Contributor

After an extension into the 2021 special session, Gov. Ralph Northam, D-Va., signed the Virginia Consumer Data Protection Act into law March 2, 2021. In doing so, Virginia became the second state to enact comprehensive privacy legislation and the first to do so on its own initiative (California led the way in 2018. but the Legislature moved forward with the bill because they were facing a ballot initiative if they failed to do so).

The CDPA's substance is not particularly new compared to recent privacy laws. It draws heavily from the proposed Washington Privacy Act and includes components similar to the California Consumer Privacy Act.

Virginia Consumer Data Protection Act (2021)

- Slated to go into effect January 1, 2023
- “The bill applies to all persons that conduct business in the Commonwealth and either (i) control or process personal data of at least 100,000 consumers or (ii) derive over 50 percent of gross revenue from the sale of personal data and control or process personal data of at least 25,000 consumers.”
- “The bill grants **consumer rights to access, correct, delete, and obtain a copy of personal data and to opt out of the processing of personal data** for purposes of targeted advertising, the sale of personal data, or profiling of the consumer.”

Tools of FTC in US

- Unfair practices
  - Injure consumer
  - Violate established policy
  - Unethical
- Deceptive practices
  - Mislead consumer
  - Differ from reasonable consumer expectations


ftc.gov
FEDERAL TRADE COMMISSION
PROTECTING AMERICA'S CONSUMERS
MAIN MENU
SEARCH
Digital Advertising Company Settles FTC Charges It Deceptively Tracked Consumers Both Online and Through Their Mobile Devices
Settlement ensures consumers can control targeted ads
FOR RELEASE
December 20, 2016