It technology and society – Zusammenfassung

Kapitel 1: Future of Work

1.1 Productivity Paradox:

Productivity growth: Ratio of output to (units of) inputs in production process, e.g., labor productivity growth

Hypothesis:

- ➔ Increase in productivity growth with rapid development in information technology (first questioned in 70s, 80s, in the U.S.)

Observation:

- ➔ Opposite finding → Slowdown in productivity growth

Reasons:

- ➔ Measurement of outputs and inputs is difficult
- ➔ Lagged impact of IT due to learning and adjustment
- ➔ Profits may be subject to redistribution and dissipation
- ➔ Inadequate management of IT

Findings in recent studies:

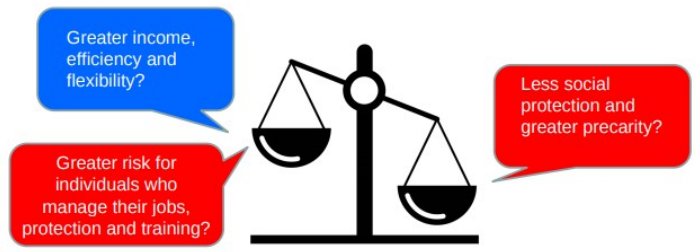
- ➔ Some evidence of differential (positive) productivity growth in IT-intensive manufacturing industries
- ➔ Depends on measure of IT intensity (automation vs. computerization)
- ➔ Never observable after late 1990s
- ➔ Positive growth driven by declining relative output accompanied by even more rapid decline in employment.

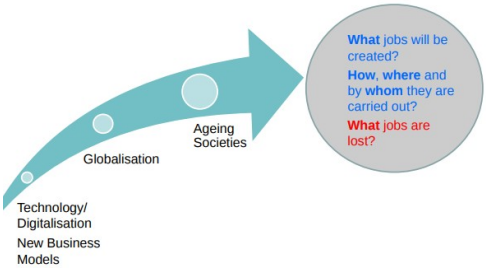
Germany:

- ➔ General decline of wages for young, medium-skilled and low-skilled workers.
- ➔ Highly-educated workers have faced declining employment opportunities in top-paying jobs
- ➔ Analytical jobs have a more positive history overall

Future:

Many pessimistic predicitons

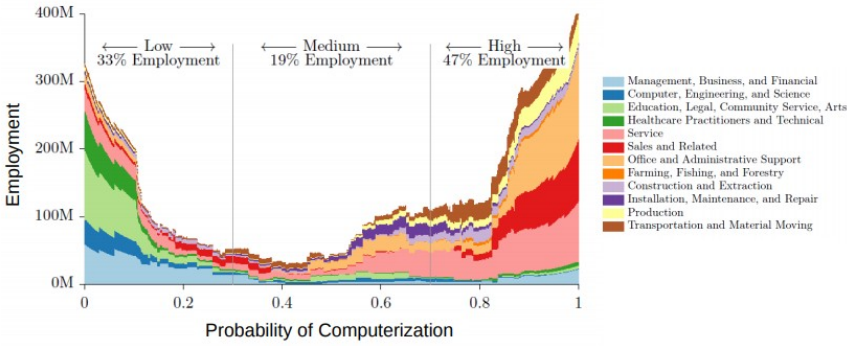




1.2 Fate of jobs (Research study):

Trends:

- ➔ Labor market polarization
  - ➔ Growing employment in high income cognitive jobs
  - ➔ Growing employment in low-income manual service occupations
  - ➔ Loss of employment in middle-income routine jobs



Position	Probability	Occupation
32	0.0065	Computer Systems Analysts
69	0.015	Computer and Information Research Scientists
109	0.03	Network and Computer Systems Administrators
110	0.03	Database Administrators
118	0.035	Computer and Information Systems Managers
130	0.042	Software Developers, Applications
181	0.13	Software Developers, Systems Software
208	0.21	Information Security Analysts, Web Developers, and Computer Network Architects
214	0.22	Computer Hardware Engineers
428	0.78	Computer Operators
442	0.81	Word Processors and Typists
…		
…		
702	0.99	Telemarketers

“The crucial problem isn’t creating new jobs. The crucial problem is creating new jobs that humans perform better than algorithms. Consequently, by 2050 a new class of people might emerge – the useless class. People who are not just unemployed, but unemployable.”

1.3 Takeaways:

- ➔ Minefield of different opinions over the last 100 years.
- ➔ It productivity is generally difficult to measure.
- ➔ Labor market polarization denotes a loss of employment in middle-income routine jobs; breeding inequality
- ➔ New business models may provide new employment opportunities but also potential societal tensions
- ➔ Need to avoid “unemployability”

Kapitel 2: Privacy

Evolving Business Models:

- ➔ Providing many more or less useful services
- ➔ Often digital content generated by users or professionally
- ➔ Improved by or even based on user data
- ➔ → Personalization/ Customization vs. Your data is the product

Data is impactful beyond online business:

- ➔ Advances in Data Analytics
- ➔ Example Healthcare: precision medicine, monitoring devices
- ➔ → requires detailed health data of many individuals

What is privacy?

- ➔ Ownership of and control over personal data
- ➔ personal dignity
- ➔ freedom to develop

Right to be left alone

New technological innovations and business models

Invasion into privacy beyond territory

No one should be able to invade your privacy by publishing information about you without your explicit consent

Exceptions: When you have published that information yourself, when in public

Boundary Regulation: Temporal dynamic process of interpersonal boundary negotiation

Negotiation of accessibility and inaccessibility that characterizes social relationships

Regulation both by how physical spaces are built and through the behaviors that take place in them

Practices are applied to achieve contextually desirable degrees of social interaction

Enormous Economic Pressure on Privacy:

“Personal data is the new oil of the internet and the new concurrency of the digital world”

Moving toward the U.S. standard of reasonable privacy expectations

Subjective: Person asserting that a search was conducted must show that they kept the evidence in a manner designed to ensure its privacy

Objective: Would society at large deem a person’s expectation of privacy to be reasonable?

Bad Privacy Examples:

Online Tracking, Cross-Device Tracking, Addition of Data from Offline World

User Protection:

- ➔ 78% state they actively protect themselves
- ➔ Browser settings (cookies): 46%
- ➔ Anti tracking software: 18%

What 3 factors shape Behavior?

- ➔ Incomplete or asymmetric information: Lack of understanding of situation
- ➔ Bounded rationality: Analysis of privacy consequences is too difficult
- ➔ Psychological aspects: e.g.: Total immersion in activity leads to lack of metacognitive monitoring

Obstacles:

- ➔ Decision making over time: Actions now have consequences later
- ➔ Choices are not and should not be perceived independent
- ➔ Research Approach-Interdependent Privacy
- ➔ Quantify the monetary value app users place on friends' personal profiles on SNS.
- ➔ Survey constructs to develop behavioral model to explain valuations.

Interpretation:

Data collection contexts affect how users value their friend's information

Sharing anonymity plays an important role in interdependent privacy valuations.

Solution Approaches:

Notice and consent

Privacy Tools

Laws and penalties

Privacy by design

Abstinence/Change Behavior

Takeaways:

Data enables many new business models; data analytics may lead to important new insights in many

Collection and monetization of data is pursued aggressively

Individuals' understanding inadequate to "solve" privacy challenges

Discussion of merits of different solution approaches needed

Kapitel 3:Privacy and Organizations

General Data Protection Regualtion:

- Regulation on data protection and privacy for all individuals in the EU and the EEA and transport of data outside these areas.
- GDPR is a regualtion not a directive.

Core Concept of GDPR: (Data Protection)

Data protection is:

- A separate right from Privacy under EU Charter
- Not rooted in concept of harm but in concept of right to control data about oneself
- Informational self-determination

Selected Definitions:

Personal data: any information relating to an identified or identifiable data subject

Data subject: a natural living person

Controller: the owner of the data

Processor: acts on behalf of the controller

Special categories: collection, storage, disclosure, transfer, profiling etc.

Automated individual decision-making: Making a decision solely by automated means without any human involvement

Profiling: Automated processing of personal data to evaluate certain things about an individual.

→ can be a part of an automated decision making process.

Data Protection Officer

Mandatory when:

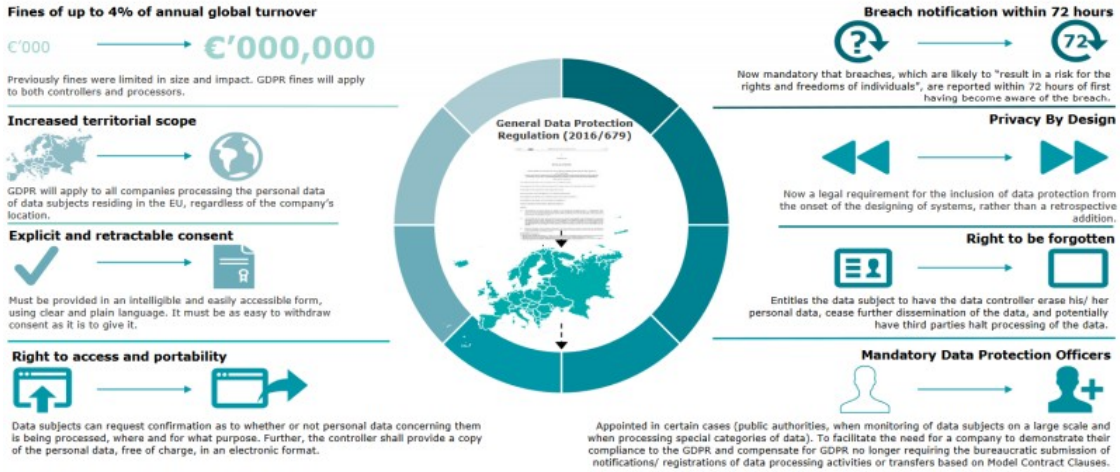
- Public authority or body
- Core activities consists of operations that require regular and systematic monitoring of data subjects
- Core activities require processing of special categories of personal data on a large scale.

Position Description:

- Involved in all issues relating to the GDPR
- Necessary resources to perform his tasks
- Independent; cannot be dismissed for exercising his tasks
- Bound by secrecy and confidentiality, liasies with management&data subjects
- Advises management and employees
- Monitors compliance with the GDPR

- Cooperates with Data Protection Authority

GDPR-Details:



How is compliance to the GDPR evolving?

Just before GDPR effective date: only 38% of businesses and 44% of charities heard of it in uk

One year after: Some news sites still blocking EU users,

Technical changes:

Product team compliance: GUI changes, back-end data logging, security, privacy by design, new tools for access

Legal team compliance: data impact assessments, internal-record breaking, renegotiating contracts, changing terms of service

What is Privacy by design?

1. Proactive not Reactive: Preventative, not Remedial;

2. Privacy as the Default setting;

3. Privacy Embedded into Design (early on in the process);

4. Full Functionality: Positive-Sum, not Zero-Sum;

5. End-to-End Security: Full Lifecycle Protection;

6. Visibility and Transparency: Keep it Open;

7. Respect for User Privacy: Keep it User-Centric.

Takeaways:

- GDPR is a behemoth: We have not even discussed Germany-specifics → Will consumers benefit
- Other competing regulations: ePrivacy Directive, Cookies and Opt-out

Kapitel 4: Privacy and Society

How Privacy Laws Conceived: (California Assembly Bill 375)

Alastair Mactaggart:

Real estate owner who became the most important privacy activist in America. Motivation: Dinner conversation with Google Engineer. Invested 2 Million Dollar in campaign

The Process:

- Idea: Gathered signatures for a statewide ballot initiative
- Approach: Hired small staff, set them up in a two-room office in Oakland and began calling privacy experts to figure what privacy campaign should say→ hiring Ashkan Soltani
- Reaction: Privacy Experts didnt take him too seriously
- Submitted final ballot initiative language to CA state
- Subsequently contacted by google, facebook
- Effort to crush ballot initiative started in the background
- → Lobbying group: Committee to Protect California Jobs
- Battle over public opinion begins

→ Submitted more than 629.000 signatures to qualify Mactaggart's initiative for the ballot

Next steps:

Compromise bill emerged that was acceptable to Mactaggart

Mactaggart offered Silicon Valley a take it or leave it privacy policy the same kind that Silicon Valley usually offered everyone else.

The Result:

AB 375 unanimously passed the assembly as the California Consumer Privacy Act (Active 2020)

What rights are included:

- What information businesses collect
- Request that information be deleted
- Get access to information on the types of companies their data has been sold to
- Direct businesses to stop selling information to third parties

Some differences to GDPR

Prevents businesses form denying service to consumers if they opt out having their data tracked and stored

But "Spotify" exception:

- Allows companies to offer different services or rates to consumers based on the information they provide
- Must be reasonably related to the value provided to the consumer by the consumer's data

Critical Opinion:

The industry's argument is that it would be too difficult for businesses to track which third parties have access to the data. If they're sharing data with third parties, they might want to have a mechanism to keep track of who they're sharing data with.

Discussion:

Complex entanglement of lawmaking

Combination of parliamentary lawmaking and ballotmeasure can lead to interesting incentive constellations, citizen got mobilized through signature collection.

Ballot measures can also be problematic e.g.: Brexit

Privacy Activism in Europe

Max Schrems:

- founded NOYB
- focuses on GDPR enforcement
- process in Europe is also influence by lobbying

LobbyPlag Project:

Comparisons of Amendments(submitted by the members of the European Parliament) and Lobby Proposals(created by the tech industry).

Democratizing Social Media:
Facebook wants social media supreme court

Process:

1. Facebook proposes change to its policy documents
2. Facebook asks users to comment on the new policy proposals (7000 user comments required, 30 days)
3. User either accept or reject proposed policies (30% of users needed to participate for the vote to be binding)

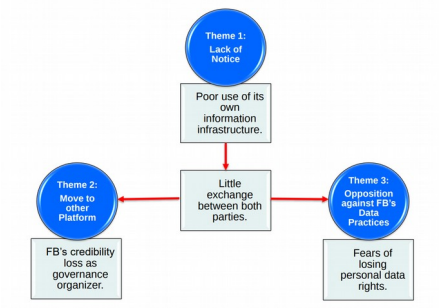
Outcomes:

None of the 3 votes reached the pre-specified participation threshold to become a binding outcome.

→ Majority rejects proposal

Problems:

1. Facebook designed and implemented a complex multilevel governance system
2. hurdles too high for user influence: 7000 comments, 30% of fb users
3. Acceptance and rejection only of entire policy document possible - not individual terms.



Takeaways:

Development of law is a messy and haphazard process

Privacy activism plays a critical role in various ways to make data protection a reality

Big questions:

How to motivate broader public to participate?

Is installing various privacy-enhancing technologies or filing a GDPR complaint more productive?

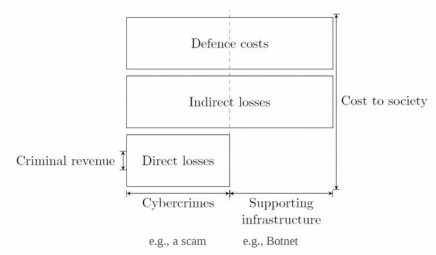
Kapitel 5: Cybercrime

What is a Cybercrime?

1. Traditional forms of crime such as fraud or forgery of forgery, though committed over electronic communication networks and information systems
2. Publication of illegal content over electronic media
3. Crimes unique of electronic networks, (e.g.: attacks against information systems, ddos attacks and hacking)

Criminal Revenue

- Revenue from crime
  - Criminal earnings – criminal inputs (investments)



Direct Losses

Value of losses, damage, or other suffering felt by the victims as a consequence of a cybercrime

Indirect Losses

Value of the losses and opportunity costs imposed on society by the fact that a certain type of cybercrime is carried out

→ Cannot easily be attributed to individual perpetrators or victims

Defense costs

Costs of prevention efforts

- Security products
- Security services provided to individuals
- Security services provided to industry
- Law enforcement

Costs > Cybercriminal Revenue

- Criminal revenue is typically significantly lower than direct losses and much lower than total losses

Policy Question

Who will pay to stop cybercrime?

- Companies: Investments in protection to reduce direct costs?
- Gonvernments: Investments in collective defense?
- Citizen: Vigilance & avoidance of “unsafe” activities, and individual protection efforts

Law Enforcement Challenges

- Perpretators and victims are often in different jurisdictions
    - Reducing the motivation and the opportunity for police action
- Mutual legal assistance across borders was not intended for routine police and criminal justice cooperation but for rare and serious cross-border crimes

Frauds:

1. Online Card Fraud

“Bellwether of online property crime overall”

Big picture:

- Payment fraud has about doubled in total/absolute value since 2012, but it has fallen slightly as a percentage of turnover

Electronic payment systems have gotten much bigger, and slightly more effcient

Online Banking Fraud: extreme growth in the last years

→ EMV → Almost all counterfeit card fraud against European cards is now at terminals outside the EU

2. Ransomware and Cryptocrime

- Ransomware exists for over 10 years: Niche crime at first
- Interesting interaction effect: After ransomware malware authors adapted to cryptocurrencies, their revenues increased substantially
- Pury cryptocurrency crimes:
    - Ponzi schemes with new emerging cryptocurrenicies
    - Crypto-mining malware
    - Fraud

3. Fraudulent Marketing and Distribution:

- Ad Fraud

- Unlicensed and patent-infringing pharmaceuticals
- Coupon and loyalty-program fraud
- Travel fraud
- Copyright-infringing software
- Copyright-infringing music and video

4. Fake Antivirus and Tech Support Scams

- Being scared into purchasing software that at best does nothing and at worst leaves your computer open to other attacks
- Complemented by so-called ‘tech support’ scams involve telephone

5. Compromised Email Accounts

- Some large-scale breaches: e.g., Yahoo’s loss of 3 Billion accounts
- Exploiting accounts being done as part of diverse set of scams

6. Fake Escrow and Other Fake Companies

Escrow:

- Third party holds and regulates payment of funds when two parties or more are involved in a given transaction
- Large variety of scams

7. Advance Fee Fraud:

- Victim must pay out a small amount of money in the expectation of a large sum of money being released
- Cormac Herley: Why are scam emails so “silly”? → Try to identify must gullible people

8. Business Email Compromise

Scam Process:

- Fraudulent email message being sent to a company’s financial manager, comptroller, or someone else with authority to execute wire transfers
- Email falsely claims to be from the CEO or other person of authority within the company and instructs the receiver to initiate a wire transfer to a foreign bank account under control of the criminal

Successful because they prey on the victim’s instinct to respond quickly to a request from a person of authority in the company.

9. PABX and other Telecom-related Fraud

- Communications Fraud Con
- trol Association (CFCA) publishes data on fraud losses associated with telephony, both fixed and mobile. → Down to 30 billion (30% down)

Industrial Cyber-espionage and Extortion

Conjecture: Enormous problem

- Criminal cases in court
- impact hard to measure

10. Fiscal Fraud

- Tax fraud and welfare fraud
  - Computer crime under the EU definition, as almost all tax returns and welfare claims are now online in many countries
- Third-party tax fraud
  - Criminals impersonating citizens by electronically filing fraudulent tax returns
  - Billions of dollars in damages

Other frauds and scams:

→ Just about every customer transaction in every type of economic activity can be subject to fraud

Large-scale Victimization Studies

1. U.S. National Crime Victimization Study
   1. About 10% of Americans affected in 2016 (Credit card and bank account fraud)
   2. About half were contacted by their institution about suspicious activity;
   3. Only about a quarter knew how the compromise occurred
   4. Only 7% dealt with police
   5. Only 12% ended suffering losses
2. UK Victimization Study by Office of National Statistics
   1. Only added cybercrime component in 2015
   2. Inclusion of fraud and computer offences has increased the total from about 6 million offences to about 11 million offences

Kapitel 6: Security in Organizations

Types of attackers:

1. Cyber criminals pursuing monetary objectives through fraud or from the sale of valuable information
2. Industrial competitors and foreign intelligence services interested in gaining an economic advantage for their companies or countries
3. Hackers, who find interfering with computer systems an enjoyable challenge
4. Hacktivists, who wish to attack companies for political or ideological motives.
5. Employees, or those who have legitimate access, either by accidental or deliberate misuse.

Evolving Attacks:

Attackers in the past:

- Broad attacks designed for mischief

Attackers today:

- Advanced attacks to acquire valuable data from an organization
- Targeted attacks against persons, organizations
- Often conducted across multiple vectors and stages
- Specialized teams using sophisticated tools & techniques
- Traditional security measures are not sufficient

Trends:

Attacks move faster → defense does not

Attacks are more targeted, upgraded techniques, number of attacks increases each year

The Evolution of Security and Risk Management

→ Concept of security and risk management has evolved as information technology has evolved

→ Scope of security and risk management has now become much larger

Setting the Scene: Complexity

- Vertical data-driven collaboration: from sensors into the cloud
- Horizontal data-driven collaboration: cross-domain, inter-organizational

Networks vs. Security

- Value of a networked system emerges from interconnection: positive network effects
- Interference from outside/inside attackers diminishes value and may lead to losses

Technology to manage these trade-offs:

- Need arises to manage security within an organization
- But also beyond organizational boundaries Coordination between different orgs.

Security Management in Theory

Motivation: Information Security Management

- Business process impose increasing requirements on information communication technology
- Increasing complexity of business processes and supporting ICT
- Necessary alignment of ICT and business processes
- Strategic protection of business assets
- Increasing set of performance regulations
- Systematic and process-oriented approach to allow for continuous improvement

→ Ensure international competitiveness

→ Meet external regualtions

→ Select, implement, and monitor efficient security controls

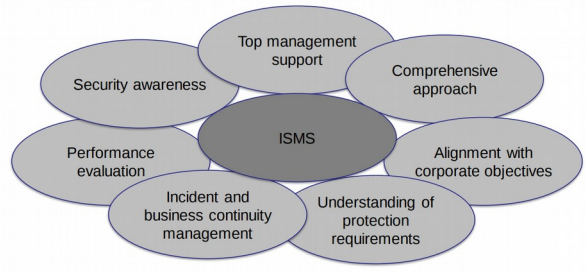
Information Security Management System (ISMS)

- consists of policies, procedures, guidelines, and associated resources
- allows an organization
  - Satisfy information security requirements of customers
  - Improve an organization's plans and activities
  - Meet the organization's information security objectives
  - Comply with regulations, legislation, and industry mandates
  - Manage information assets

Benefit of Standards for Information Security:

- Cost reduction
- Appropiate security level → increased comparability
- Competitive advantage → certificate of enterprise or products

ISMS Success Factors



- Built-in performance approach

Summary of Theory:

- Risk-driven information security management: Selection and implementation
- Continuous improvement should correspond with changing (risk) environment
- Documentation-centered approach

- Applicability: Generic, adoptable, flexible, expert know-how necessary
- Certification schema available and increasingly used in practice

Challenge 1: Decisions should be based on Reliable Data → Ransomware reaches all time high

Challenge 2: Do not reinvent the wheel. → Attacks on Sony and Swift

Challenge 3: Security Standards should be based on Scientific Evidence → Passwords

Challenge 4: We should act quickly! → It usually takes too long to detect an attack

Summary Challenges:

- Many high-quality IT security research projects, but too little data-driven collaboration with industry and policy actors → We often do not know effective security management is in practice but public data does not show a pretty picture
- Science of security management: Scientifically validated approach to prioritize security measures is missing, but needed

Takeaways:

- External and internal drivers push for a systematic information security management → Strategic protection assets, compliance regulations, growing complexity
- Information security management systems (ISMS)
  - Meet security objectives, satisfy external requirements & regulations, improve security-related activities…
  - Established standards available to help, but practice is mess

Lecture 7: Security issues in society

United States/Five Eyes:

- Efforts since the early 2000's to "master the Internet" by the NSA and Five Eyes
- Our knowledge about diverse efforts has improved: Ed Snowden and other whistleblowers leaked information about the capabilities and methods of Western Intelligence services

Programs

1. Prism: NSA codename for an access channel that had been provided to the FBI to conduct warranted wiretap
2. Telecommunications: AT&T provided the NSA with access to billions of communications records including emails and phone call data + surveillance systems in internet hubs
3. Tempora: To collect intelligence from international fiber-optic cables → use of 10000s of selectors to sift through 600 Million "telephone events"

4. Muscular: Collection of data as it flowed between the data centers of large service firms such as Yahoo and Google → data often flowed unencrypted in the clear
5. Special Collection Service: Various strategies, e.g., to implant collection equipment in foreign telecommunications providers, Internet exchanges and government facilities
6. Longhaul, Quantum: Focus on encrypted communications
7. Xkeyscore: Distributed database enabling an analyst to search collected data remotely and assemble the results
8. Hacking: Computer and network exploitation→tools of NSA are currently fairly well understood

Vision/Goal to be adopted by governments:

Role of private Parties:

Developers of malware to be used by nation states against terrorism etc.

<u>People → Nothing to hide argument</u>

The argument that no privacy problem exists if a person has nothing to hide is frequently made

→ People believe there is no threat to privacy unless the government finds unlawful activity

→ Security is the strongest argument for government surveillance

<u>Important Opposing View</u>

- Nothing-to-hide argument is an anti-social statement
- Privacy should not be treated as something to hide, but something to protect
- Balancing privacy against security? Or Balancing freedom against control?
- “I do not care what happens, so long as it does not happen to me!”

<u>Impact of Filtering:</u>

- Placing restriction freedom of speech in a non-transparent way

<u>Takeaways:</u>

- Significant efforts being done towards implementing data collection and processing in the name of national security

- Similar efforts undertaken in many countries regarding Internet filtering and censorship
- Many challenges which are hard to resolve:
  - Grand challenges: Impact on civil liberties
  - More well-defined challenges: Like responsible disclosure

Lecture 8: Behavioral Insights and Societal-scale Mechanisms

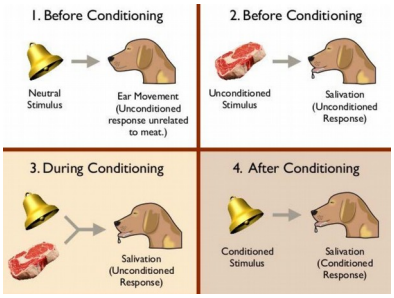
From Data to insights:

- Big data provides wide access to unprecedented amounts of data, offers new insights into human emotions, cognitions, motivations, decisions, preferences, behaviors, and interactions, and facilitates the data-driven development of new conceptual ideas in different fields. → Unemployment statistics compared to google searches “looking for a job”

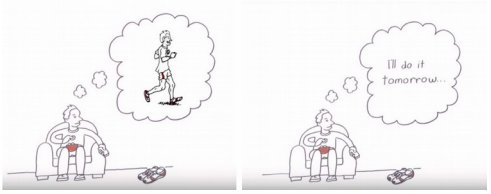
Behavioral Insights:

- The behavior of the individual has been shaped according to relevations of “good conduct” never as the result of experimental study.

Stimulus response:



Behavioral Insights: help us to understand how people actually make decisions in everyday life.



Behavioral Insights:

- An inductive approach to policy making that combines insights from psychology, cognitive science, and social science with empirically-tested results to discover how humans actually make choices.

Nudge theory:

→ Richard Thaler (“The father of the nudge theory” Nobel economics prize winner in 2017 for his contributions to behavioral economics)

Don’t push. Don’t pull. “Nudge”.

1. A nudge, as we will use the terms, is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives.
2. Nudges are ways of influencing choice without limiting the choice set or making alternatives appreciably more constly in terms of time, trouble, social sanctions and so forth. They are called for because of flaws in individual decision-making, and they work by making use of those flaws.

- To encourage people to make decisions that are in their broad self-interest through a relatively subtle policy shift.

- It is not about penalising people financially if they don't act in a certain way.
- It is about making it easier for them to make a certain decision.
- Nudges are specifically designed to preserve full freedom of choice.

Digital Nudging

- The use of user-interface design elements to guide people's behavior in digital choice environments.
- Digital choice environments are user interfaces, for example web-based forms.

Types of Nudges:

1. Default Option
2. Social Proof Heuristics
3. Reminder
4. Providing Feedback
5. The Element of Entertainment
6. Disclosure

1. Default Option:

A default option is simply what happens if you do nothing. An individual is nudges to choose a given option if it is set as default.

Examples: Default setting to "tipping"

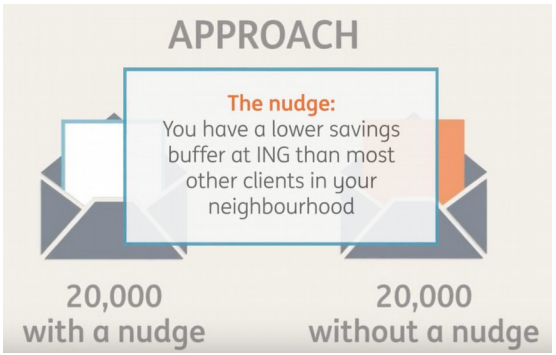
2. Social Proof Heuristics (Social Norms)

Injunctive Norms: behavior other individuals approve of. (80% of individuals think activity x is morally good)

Descriptive Norms: the desirable behavior of others. (80% of individuals engage in desirable activity x)

Such information is often most powerful when it is as local and specific as possible (the overwhelming majority of people in your community do x)

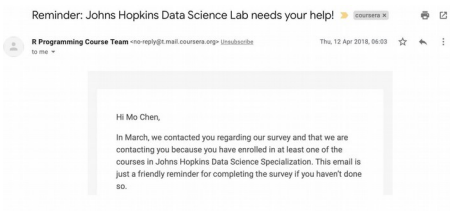
→ Bank Saving



3. Reminder

For example, by email or text message, as for overdue bills and coming obligations or appointments

→ E-learning



4. Providing Feedback

The feedback makes people aware of their behavior and pushes them into the desired direction.

Receiving positive feedback gives a good feeling and serves as a reinforcer.

→ Speeding measures

→ E-health

5. The Element of Entertainment/Gamification

Humans have a need to integrate play elements into their lives.

To nudge is to stimulate the desired behavior in an entertaining way.

→ Snapchat Throphies → E health etc.

6. Disclosure

In some settings, disclosure can operate as a check on private or public inattention, negligence, incompetence, wrongdoing, and corruption.

→ Smart Meters → E-govermment

China's Social Credit System:

- "is a real life black mirror nightmare", "big data meets big brother as china moves to rate its citizens"

What is it?

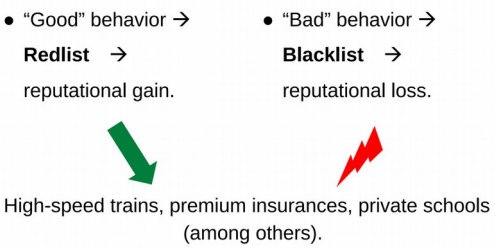
- A multi-level nationwide rating system.
- A core concept of the SCS: honest and trustworthy
- All legal entities receive 18-digit ID

→ "Unified Social Credit Code"

→ Allowing the trustworthy to roam everywhere under heaven while making it hard for the discredited to take a single step

Reasons:

1. Moral decline in Chinese society
    1. Public shaming and praising.
2. Strengthening the domestic economy.
    1. Loans based on trustworthiness rather than financial worthiness.



Levels: National Level, Municipal/City Level, Commercial Level

Findings: SCS at the Current Stage:

- Focus on public shaming & praising.
    - Blacklists: clear punishments.
    - Redlists: vague rewards.
- Currently: more information on "bad" behavior than on "good" behavior.

Takeaways:

- Behavioral insights is widely used in various fields and in many countries
- Behavioral insights is becoming increasingly popular in digital environment to shape our behaviors and our society
- The unprecedented scale of these mechanism necessitates careful study and involvement of citizens to avoid (likely) impacts such as surveillance, oppression etc.

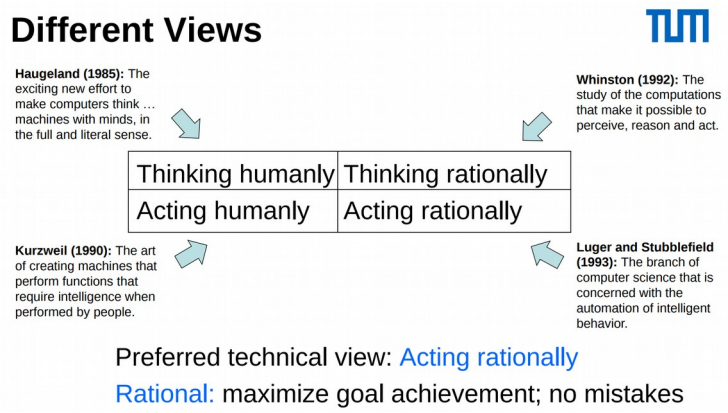
Lecture 9: Introduction to Artificial Intelligence

What is Artificial Intelligence?

- No clear consensus on the definition of AI
- John McCarthy coined the phrase AI in 1956

  → It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human or other intelligence, but AI does not have to confine itself to methods that are biologically observable.

  → Intelligence is the the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.



Another "Working Definition" of AI

Artificial intelligence is the study of how to make computers do things that people are better at or would be better at if:

- They could extend what they do to a World Wide Web-sized amount of data
- Not make mistakes

A note about robotics:

- Density of robots in industrial facilities: ø74 per 10.000 employees
- China below average with 68, Germany at 309
- 31 Million robots in households → rapidly growing

How to Measure "Success" in an AI world?

→ Turing Test

→ Loebner Prize (since 1991)

Strong AI vs Weak AI

- Strong AI is artificial intelligence that matches or exceeds human intelligence – the intelligence of a machine that can successfully perform any intellectual task that a human being can.
    - Primary goal of artificial intelligence research and an important topic for science fiction writers and futurists
    - also referred to as “artificial general intelligence” or as the ability to perform “general intelligence action”
- Weak AI is an artificial intelligence system which is not intended to match or exceed the capabilities of human beings, as opposed to string AI, which is. Also known as applied AI or narrow AI.

How intelligent or conscious can machines get?

→ The Chinese room experiment: Suggests that a computer can never understand what it does, because – like you- it just executes the instructions of a software program. → Even if a machine seems intelligent, it will never be really intelligent.

Fears related to AI:

1. Impact on the job market → is AI primarily job-replacing or job-enabling?
2. Opposing views about a glorious AI-centric future versus a dystopian AI-dominated future

AI for Good foundation: how can AI solve society’s biggest challenges?

1. Sustainable AI: Food, energy and water
2. Environment and AI: healthy oceans, protect wildlife
3. Health and AI: health, sleep, nutrition
4. Transparency and AI: Fighting corruption
5. Education and AI: Personalized education

also: harvesting human intelligence as a byproduct of fighting artificial intelligence advances

→ Captcha → reCaptcha (completely automated public Turin test to tell computers and humans apart

Case Study: Stack Overflow

- The good: Ready-to-use code examples, and 78% of software developers look up programming questions on Stack Overflow on a daily basis
- The Bad: 30% of code examples on Stack Overflow contain software security vulnerabilities
- The ugly: Such bad code examples were detected in over 190.000 apps available on Google Play

Dissecting the Problem

Insecure suggestions had on average: higher view counts, higher scores, more duplicated than secure suggestions

Solution Approaches:

- Do not simply reuse code from the web
- Install and use code analysis tools

- Use simplified APIs
- Mandate usage of documentation
- Warnings: Security model to identify insecure code and inform the user
- Recommendations: Use case and similarity model to provide a way out
- Reminder: If users copy insecure code examples to the clipboard, warn and recommend again
- Defaults: Up-rank secure code examples in the search results

Takeaways:

- Artificial intelligence research, practice and deployment has a long and rich history
- Very different directions: modeling human intelligence versus creating actionable systems delivering output that is difficult to accomplish for humans
- Stack Overflow case an example for AI for Good: However, also raises many questions about deployment

Lecture 10: moral machines & the ethical dimensions of artificial intelligence

1. Introduction to ethics/moral philosophy:

1. Utilitarianism
2. Deontology
3. Virtue Ethics

2. Digital ethical dilemma:

1. Autonomous driving

   → How do moral principles cope with morally-charges scenarios caused by digital technology?
   → In any algorithmic system, a morally-charged decision process can be guided and evaluated by utilitarian, deontological or virtue ethics principles.

Ethics: Key question: "What is the right thing to do?"

→ Scenarios: Are numerous lives worth more than a single one

Moral principles from these examples:

1. Consequentialist principles (utilitarianism)
   1. locates morality in the consequence of an act → What created the most overall utility for the individuals involved?
2. Categorical principles (deontology)
   1. locates morality in certain absolute/universal moral duties and rights regardless of the consequences. → What is the intrinsic quality of the act itself?

Origins of utilitarianism (Jeremy Bentham)

“Nature has places mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do.”

Fundamentals:

- Moral justification of the Homo oeconomicus
  - Maximize beneficial outcome(pleasure), minimize detrimental outcome(pain)
  - Pragmatic: finite resources, best outcome for all parties involved.
  - Most constitutions in the Anglo-Saxon world modelled after utilitarian priciples (United Kingdom, Usa, Australia)

Origins of deontology (Immanual Kant)

“act only according to that maxim whereby you can at the same time will that it should become a universal law without contradiction”

Fundamentals:

- Moral reasoning and actions based on ideals.
  - Categorial imperative:
    - Moral decision making on the basis of duty whereby an individual must not be instrumentalised to serve the betterment of another person’s condition.
- Adopted by the German constitution(article 1)

How should machines make moral decisions?

→ Global study on moral reasoning in autonomous driving

“What are the values that we will embed in the car when it makes life-and-death choices?”

- About the study
  - 40 milion decisions recorded.
  - From 233 countries and territories
  - 492,921 subjects completed the optional demographic survey at the end
- Results
  - Individual variations
  - Cultural clusters
  - Country-level predictors

Nine factors influence moral reasoning:

1. Sparing humans vs pets.
2. Staying on course vs swerving.
3. Sparing passengers vs. pedestrians.

4. Sparing more lives vs. fewer lives.
5. Sparing men vs. woman
6. Sparing young vs. the old
7. Sparing pedestrians who cross legally vs jaywalking.
8. Sparing the fit vs. the less fit
9. Sparing those with higher vs. those with lower status

Results of the study:

Important to less important: Stroller, girl, boy, pregnant, doctors, athletes, executives, large people, old people, pets, criminals

→ Cultural Clusters:

- Cultural proximity results in converging preferences for machine ethics (in this context)
- Do between-cluster differences pose greater challenges? → Yes

→ Cultural differences:

- Eastern: less preference for the young over the old.
- Southern: more preference for the young over the old.
- Southern: more preference for higher status
- All: weak preference for sparing pedestrians over passengers.
- All: moderate preference for sparing the lawful over the unlawful

<u>Striking cultural differences (Southern cluster)</u>

1. men > women 2. fit > large 3. status>poverty

→ Implications for manufacturers of autonomous cars? Should they implement these preferences?

- Everyone would be better off if Avs were utilitarian (in the sense of minimizing the number of casualties on the road), but they all have a personal incentive to ride in Avs that will protect them at all costs.

<u>Is virtue signaling the way out of this social dilemma?</u>

- Emphasizes the virtues, or moral character, in contrast to the approach that emphasizes duties or rules (deontology) or that emphasizes the consequences of actions (consequentialism)

How do countries usually make (legal) decisions about ethicallly-charged technologies?
→ Ethics commisions: a group of ethicists and other experts (hired by the government).

→ Humans over pets, more over fewer, young over old, high over lower status, lawful over unlawful

Takeaways:

Three major approaches to normative ethics:

1. Deontology (duty, rule based)
2. Utilitariasm(consequences, outcome)
3. Virtue ethics(moral character)

Potentially many digital technologies create morally-charged scenarios.

Different cultures may have different ethical preferences or should experts make decisions?

Autonomous vehicles create ethical and social dilemma?

Lecture 11: Fairness, Accountability & Transparency

Fairness, Accountability & Transparency (FAT) of AI Algorithmic Systems in general

Across multiple other task domains, every single decision made by an algorithm involves an ethical dimension:

- Predictive policing and jurisdiction (recidivism)
- Predicting financial worthiness (credit scoring)
- Predicting employees' success (hiring decisions)

Line of argumentation:

- Is the decision fair?
- Who made the decision? → Who is responsible? Where rests accountability?
- How was the decision made? → Can we understand the decision-process? How transparent is the process?

→ 78% of Americans do not trust Avs

FAT: Trust-enhancing factors for AI adoption:

Without transparency, can we know whether the decision was fair or who is responsible for it?

Is transparency a necessary (and sufficient?) condition to determine accountability and fairness in an algorithmic system?

Case-study: Assessment tools to predict recidivism risk

- How likely is a defendant to commit a felony once released from prison?

Appeal of risk assessment tools:

- U.S. locks up far more people than any other country, a disproportionate number of them black.
- Key decisions in the legal process have been in the hands of human beings guided by their instincts and personal biases.
- If computers could accurately predict which defendants were likely to commit new crimes, the criminal justice system could be fairer

→ Eric Loomis sentenced to 6 years in prison because classified as individual who is at high risk to the community by COMPAS software tool

Is "COMPAS" fair?

How does the algorithm calculate the score?

- COMPAS in use since 2000 (predictions for > 1 million offenders)
- Scores from 1-10 (10 = highest score)
- Algorithm is proprietary and thus a trade secret: → Little transparency over decision-making process.

ProPublica study:

- Analysis of COMPAS risk score of 7000 people arrested in Florida in 2013 and 2014
- Only 20 percent of the people predicated to commit violent crimes actually went on to do so.
- For misdemeanors, such as driving with an expired license, the algorithm was just above 50 percent correct.
- Overall: of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

→ Propublica says COMPAS is not fair → conceptualize fairness from the perspective of the defendant

→ Northpoint says COMPAS is fair. → conceptualize fairness form the perspective of the sentencer

Northpoint's fairness definition:

Medium to high risk scores map to equal probability in actual re-offending among both black and whites.

- Prediciton: black/white person → risk score 7 (medium)
- Reality: among both black/white with risk score 7 equal rate of re-offending (f. ex. 60%)

→ Advantage: Judges do not need to consider race at all! = Northpoint's definition of fairness.

ProPublica's fairness definition:

- Problem: there are more black people in the training set that re-offend.
- Blacks twice as likely to be classified as medium or high risk (42% vs 22%)

- Northpoint is interested in the set of people re-offended.
- What about the people who ultimately did not re-offend?
- Test-based predictor is unbiased by race: exactly 60% of blacks classified as “high risk” recidivate and 60% of whites. Also, exactly 30% of blacks are classified as “low risk” of whites.
- However the “high risk” group contains 60 blacks and 40 whites, while the “low risk” group contains 40 blacks and 60 whites.

	White		Black		All	
	Low Risk	High Risk	Low Risk	High Risk	Low Risk	High Risk
Distribution:	**60**	**40**	**40**	**60**		
No Recidivism	42	16	28	24	70%	40%
Recidivism	18	24	12	36	30%	60%

- Note: Distribution across high and low risk differs across race

  → Attributes associated with being black more likely correlated with higher risk

  → These are mostly demographic factors since race is not a feature considered by the algorithm.

  → So-called proxies: features that are correlated with class membership.

→ DATASET IS BIASED TOWARDS BLACKS, ALGORITHM IS NOT

<u>Is this a problem? What about the people who ultimately did not reoffend?</u>

False positive: She must stay in custody even though she poses no threat to society.

False positive rate for blacks: 46,15% → almost half of people classified as high risk pose no threat.

False positive rate for whites: 27,69% → almost half of the whites classified as low risk ended up committing a crime

True positive: She must stay in custody and she is a threat to society.

False negative: She can go home even though she poses a risk to society.

True negative: She can go home and poses no threat to society.

<u>ProPublica’s conclusion:</u>

- COMPAS was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.
- White defendants were mislabeled as low risk more often than black defendants.
- This is ProPublica’s conceptualization of fairness: Keep false positive and false negative rate equal.

→ Impossible to simultaneously satisfy both definitions of fairness because black defendants have a higher overall recidivism rate.

The bias in the data:

→ The broken window theory: a cycle of crime: neighborhoods with visible civil disorder → more police forces → more arrests

ProPublica study: Bias in the data

A COMPAS questionnaire created a cycle:

- COMPAS assessment is based on 137 features about an individual an the individual's past criminal record. Defendants answer 137 questions.
- Answers are fed into COMPAS software to generate recidivism risk scores.
- Race is not one of the questions → again: many proxies for race

Challenges to Ensure Fairness in Machine-learning Based Decisions:

- Garbage in/garbage out: Data can be consistently biased
- What are meaningful fairness criteria?
- How do different criteria relate and create trade-offs?
- What are their limitations?

Algorithmic systems:

- Cannot consider multiple conceptualization to fairness.
- Each definition may have its benefit and disadvantage for the data controller and data subject
- Who decides on the specific fairness definition is in position of power.

"A right to explanation" of automated decision making in the GDPR:

A right to explanation of all decisions made by automated or artificial intelligence is legally mandated since May 2018

1. Create transparency about automated decision making.(right of data subject)
2. Create accountability (duty of data controller).

→ Create trust: the comfort in making oneself vulnerable to another entity in the pursuit of some benefit.

Ex ante explanation: explanation before the decision was made → can logically address only system functionality, as the rationale of a specific decision cannot be known before the decision is made.

Ex post decision: explanation after the decision was made → occurs after an automated system has taken place. Ex post explanation can address both system functionality and the rationale of a specific decision.

What does explanation mean?

Meaningful information about the logic involved, as well as the significance and the envisaged consequences of automated-decision making. → >= 3 attributes.

Takeaways:

1. Fairness, accountability and transparency can serve as ethical measurements.
2. Fairness, accountability and transparency are trustenhacing factors → product adoption.
3. While algorithms outperform humans on a variety of tasks, they may systematically and consistently discriminate if the data contains human biases.
4. Raw data is an oxymoron!
5. Algorithmic systems can only implement one conceptualization of fairness.
6. The GDPR contains a “Right to Explanation” but only grants data subjects ex ante explanations.
7. Potentially all ML-based systems face FAT challenges if they make prediction on individuals.