

# Kickoff meeting

Proseminar: Computer und Ethik

---

Daniel Mock, Peter Rossmanith

10. Oktober 2023



Theoretical  
Computer Science

**RWTH**AACHEN  
UNIVERSITY

# Einführung

---

Computer und Informationstechnologie, so wie auch die Informatik als Wissenschaft, üben wohl unbestritten einen stets wachsenden Einfluss auf die Lebensbedingungen so gut wie aller Menschen, und damit auch auf die Gesellschaften, aus. Menschliche Arbeitskraft wird durch Maschinen ersetzt, Medien lassen sich unbegrenzt vervielfältigen, Daten aller Art zu sammeln und zu verknüpfen ist in völlig neuen Größenordnungen möglich geworden. Dies sind nur drei Beispiele für Möglichkeiten, die die Informationstechnologie mit sich bringt. Aber sie deuten schon an, dass hier Entscheidungen für den Einzelnen und die Gesellschaft zu treffen waren und sind, für die es keine historischen Vorbilder zu geben scheint.

- **Kickoff / Fragen** (Heute)
- **Themenvergabe / Fragen** (Nächste Woche)
  - Jeder bekommt ein Thema (eine Person pro Thema)
- **Vorbereitungsphase**
- **Wöchentliche Präsentationen** (Anfang: November)
  - Feedbacktreffen mit mir
  - Eine Woche vorher: Thesenpapier
  - 15-20 Minuten Präsentation + ca 45 min Minuten Diskussion
- **Gemeinsame Ausarbeitung** (bis. 15. Februar)

- Zentrale Punkte mit *Prämissen* ausarbeiten
- Sachverhalte wiedergeben (Was können Technologien aktuell/in naher Zukunft? welche Gesetze gibt es? Welche Gruppen gibt es? etc.)
- Zentrale Konflikte
- Lieber konkrete Einzelfälle anschauen statt vage und allgemein
- Fruchtbare Debatte fördern

## Beispiel: Information Paradox

- Nutzer wollen möglichst einfachen, günstigen Zugang zu Informationen
  - Wichtige Informationen sollen sicher und zuverlässig sein
- Widerspruch → Muss in *Einzelfällen* aufgelöst werden
- Ständige Einzelabwegung beim Design von Informationstechnologie

Thesenpapier:

- Vorbereitung zum Vortrag + Diskussion
- Quellen zum Weiterlesen

Vortrag:

- Nicht zu sehr ins Detail gehen
- Folien übersichtlich halten
- Vortrag nur Startpunkt: Ihr müsst mit dem Thema tief vertraut sein
- Aufteilung in zwei Teile möglich (Vortrag 1, Diskussion 1, Vortrag 2, Diskussion 2)

Diskussion:

- Moderation: Etiquette einhalten, Diskussion thematisch lenken, Argumente herausarbeiten/untermauern/widerlegen

- Startpunkt: Unsere Quellen und Anreize
- Zeitungsartikel, Paper, Bücher, Vorträge, . . . .
- Popsience als Start ist OK, nutzt aber lieber fachlichere Quellen
- Sucht weitere Quellen selbstständig heraus
- Achtet auf Ausgeglichenheit (Verschiedene Standpunkte, nicht nur deutsch etc.)
- Überprüft Fakten und Argumente
- Müsst euch nicht an unsere Quellen halten
- Eigenen Fokus finden

- Schreibt über eure Erkenntnisse während der Recherche
- Geht insb. auf die Diskussion ein
- Recherchiert die gemachten Argumente nach
- Gebt eure Quellen an
- 3-4 Seiten pro Thema
- Zusammen in eine gemeinsame Ausarbeitung
- Schreibt eine gemeinsame Einführung

- Mindestens ein persönliches Treffen mit mir vor der Ausgabe des Thesenpapiers
- Diskutieren Struktur, Kernpunkte etc. des Papiers/Vortrags und eure Fragen
- Gerne mehr Treffen!
- Terminfindung per Mail, Telefon oder bei mir im Büro

mock@cs.rwth-aachen.de

+49 241 80 21132

- Man kann sich ab jetzt 2 Wochen lang ohne Fehlversuch abmelden

## **Eigenes Thema:**

- Thesenpapier
- Vortrag
- Moderation der Diskussion
- Schriftliche Zusammenfassung

## **Wenn man nicht selber vorträgt:**

- Dem Vortrag genau zuhören
- Eigene Fragen notieren
- An der Diskussion teilnehmen
- Feedback geben

Bevor ihr beginnt:

- Versteht das Thema, Quellen recherchieren
- Gliederung erstellen, Themen/Fokus auswählen
- Gute Beispiele finden
- Mögliche Fragen und offene Themen für Diskussion finden
- Wollt ihr auf Deutsch oder Englisch präsentieren?

Während der Vorbereitung der Folien:

- Kontext geben, ggf. auf vorherige Vorträge referenzieren
- Langsam präsentieren. Nicht jeder wird euch sofort verstehen
- Aufgeräumte Folien: Lieber ein Bild als eine Textwand
- Nutzt einfache Beispiele um Ideen zu präsentieren. Gute Beispiele können verständlicher sein als ein Beweis
- Diese Folien sind viel zu voll und ein schlechtes Vorbild

- Viele verfügbar, z.B. <http://ianparberry.com/pubs/speaker.pdf>
- Learning by doing
- Lernt von den Erfolgen und Fehlern anderer
- Üben, üben, üben!

## Stanford Encyclopedia of Philosophy

- Information Technology and Moral Values
- Privacy and Information Technology
- Social Networking and Ethics
- Philosophy of Technology
- *Ethics in Cybersecurity*

<https://library.oapen.org/handle/20.500.12657/47324>

## Die Themen

---

- Was sind Hate Speech, Fake News und Propaganda?
- Welche Auswirkungen haben diese auf die Gesellschaft?
- Wie sollte öffentlicher Diskurs aussehen? Was darf/sollte wie von wem reglementiert werden? Betreiber, der Staat, die Öffentlichkeit?
- Ethics of Cybersecurity: Kapitel 11, 12 und 18
- *LikeWar: The Weaponization of Social Media*
- Absprache mit Thema “Filterblasen”

- Ab wann sind Filterblasen überhaupt problematisch?
- Oft problematisch bei Diskussionen, da verschiedene Vorstellungen von Filterblasen existieren
- Epistemische Filterblasen vs Echokammern (Nguyen)
- Dynamische komplexe Systeme
  
- Nguyen: <https://www.cambridge.org/core/journals/episteme/article/abs/echo-chambers-and-epistemic-bubbles/5D4AC3A808C538E17C50A7C09EC706F0>
- Chapter 3.5 in Social Networks and Ethics
- Social Media Curation <https://youtu.be/Ji1Eb42q-CI>

# Einfluss von “virtueller Realität” auf echte Realität

- Schönheitsideale durch Instagram, Trends von Tiktok
- Erwartungshaltung, dass man auf SNS vertreten und immer verfügbar ist
- Auswirkungen auf Freundschaften?
- Hyperrealität (Kapitel 2 von “Social Networking and Ethics”)
- Wurden ethische Werte durch SNS verändert? Unterschiedliche Werte und Verhalten online vs offline?

# Autonome und ferngesteuerte Waffensysteme

- Welche Waffensysteme sind/werden autonomisiert? Zu welchem Grad? Human in the loop?
- Aufwärtsspirale der Gewalt durch automatisierte (unbetreute) Aggression?
- Verschwimmende Grenze zwischen Realität und Spiel bzw. Abstraktion von Gewalt und Verantwortung?
- Was sind die aktuelle Haltungen von Regierungen, Militär, Thinktanks, NGOs?
- Wie ist die Nutzung in aktuellen Kriegsgebieten wie Ukraine und Berg Karabach?
- *Army of None: Autonomous Weapons and the Future of War*

- Anfangs viele Frauen, inzwischen männerdominiert
- Welche Unterschiede gibt es zwischen damals und heute?
- Im Iran viele Frauen in der Informatik. Wieso?
- Was sind die Erfahrungen an der RWTH (STEM)? Umfrage?
- Aktuelle Entwicklung?
- Ist das ein Problem? Woher kommt es? Was kann man dagegen tun?
- <https://www.derstandard.de/story/2000099056611/frauen-regierten-einst-die-informatik-dann-war-geld-im-spiel>
- <http://mod.iig.uni-freiburg.de/cms/fileadmin/publikationen/online-publikationen/Frauenanteil.Informatik.International.pdf>

- Verschiedene Modelle (in verschiedenen Kontexten anschauen)
- Scoring (Schufa, Social Credit)
- Nutzerbasierte Bepreisung
- Surge pricing
- Welchen Einfluss hat es auf die Gesellschaft?
- Wer profitiert davon?
- Ist es fair?
  
- Dynamic Pricing: <https://youtu.be/J9X055011C0>
- What Money can't buy: The Moral Limits of Markets
- <https://www.wirtschaftsdienst.eu/inhalt/jahr/2018/heft/12/beitrag/eine-vermessene-gesellschaft-braucht-transparenz.html>

- Woher kommen die Daten? Wie aussagekräftig sind sie? Haben sie Biases?
- Biases lernen? Mögliche Auswirkungen?
- Welche Daten sollte man überhaupt betrachten?
- Verantwortung mit Daten und von Ergebnissen?
- <https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf>
- <https://www.theguardian.com/technology/2019/jul/29/what-is-facial-recognition-and-how-sinister-is-it>
- <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0177-4>

- Einfache Vervielfältigung Problem für Copyright
- Heute den technischen Realitäten adäquat?
- Copyright von Input und Output von LLMs und generativer AI?
- Unterschiede in anderen Kulturen?
- Verschiedene Medien brauchen verschiedene Formen: Software, Musik, Bücher, ...

# AI Alignment und Security in LLMs

- Welche Probleme tauchen bei der Nutzung von LLMs (nur LLMs, nicht AGI) auf?
- Toxicity, Hallucinations, etc...
- Welche Möglichkeiten gibt es?
- Missbrauch von LLMs vorbeugen: Watermarking
- Bsp: LLMs mit menschlicher Interaktion (psychologische Erstberatung)?  
Transparenz?
- “Ethische” Nutzung von LLMs
- Absprache mit “Big Data und Diskriminierung”
- [https://en.wikipedia.org/wiki/AI\\_alignment](https://en.wikipedia.org/wiki/AI_alignment)
- <https://scottaaronson.blog/?p=7230>

- Wann ist Hacken ethisch?
- “Hacker culture” heute und damals
- Veröffentlichen von Sicherheitslücken, Eindringen in andere Computer Systeme
- *The Ethics of Cybersecurity* Kapitel 9, 15
- [https://en.wikipedia.org/wiki/Hacker\\_ethic](https://en.wikipedia.org/wiki/Hacker_ethic)

- XOR zum vorherigen Thema
- Was ist Cyberabwehr?
- Wie ist die Lage in der EU und Deutschland? Welche Institutionen, Initiativen und Gesetze gibt es?
- Was sollte der Staat garantieren/fördern?
- Wie ist die Gefahrenlage?
- Offensive Cyberabwehr (Hackbacks)?
- *The Ethics of Cybersecurity* Kapitel 9, 15
- <https://www.euractiv.de/section/innovation/news/nationale-sicherheitsstrategie-hackbacks-kommen-nun-doch-nicht/>
- <https://www.baks.bund.de/de/arbeitspapiere/2020/von-firewall-bis-hackback-das-spektrum-militaerischer-cyberoperationen>

# Privatsphäre vs. Strafverfolgung

- Quellen-TKÜ, Bundestrojaner
- Hintertüren
- Scan on Device (z.B. bei Apple)
- Verschlüsselung verhindert kategorisch dem Staat viele Möglichkeiten
- Abwägung zwischen Strafverfolgung und Privatsphäre
  
- Ethics of Cybersecurity: Kapitel 8
- <https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/ethical-questions-about-encryption/>
- <https://www.heise.de/news/Kinderporno-Scan-auf-dem-iPhone-Apple-haelt-am-Aus-der-Funktion-fest-9.html>

# Technologie im Blick der (Pop-)Kultur und deren Hintergründe

- Aufarbeitung Sicht der Kultur auf Technologie und Fortschritt
- Nicht einfache Wiedergabe, sondern Herausarbeiten unterliegender Themen; Hintergrund in damaliger Gesellschaft und (mögliche) Weltanschauung des Autors; Philosophische Konzepte
- Vergleich über Zeiten und Gesellschaften hinweg
- Rückblick auf 'Vorhersagen' ?
- Kultur: Bücher, Filme, Spiele, Abhandlungen ...
- Findet eine breitgefächerte Auswahl! Müsst viel lesen und schauen
- Erste Beispiele: Matrix, 2001, Wargames/Terminator 2, Deadalus und Ikarus, Prometheus, Star Trek, Solaris, Cyberpunk(Spiel, Bladerunner, Necromancer), Forbidden Planet, The Forbin Project, The Bicentennial Man, Artificial Intelligence (der Film) etc

1. Meinungsfreiheit vs. Propaganda
2. Filterblasen
3. Einfluss von "virtueller Realität" auf echte Realität
4. Autonome und ferngesteuerte Waffensysteme
5. Geschlechter in der Informatik
6. Scoring und Pricing
7. Big Data und Diskriminierung
8. Intellectual Property im digitalen Zeitalter
9. AI Alignment und Security in LLMs
10. Cybersecurity und Ethisches Hacken / Cybersecurity in Deutschland und EU
11. Privatsphäre vs. Strafverfolgung
12. Technologie im Blick der (Pop-)Kultur und deren Hintergründe

## **Wöchentlichen Termin finden**

---

Ab KW 44 (30+. Okt.), 90min Block, im Informatikzentrum

- Montag, 10:30 - 12:00
- Donnerstag, 12:30 - 14:00
- Donnerstag, 14:30 - 16:00

Standardtermin: Donnerstag, 12:30

# Wintersemester 2023/24

Oktober	November	Dezember	Januar	Februar	März
1 So	1 Mi Allerheiligen	1 Fr	1 Mo Neujahr	1 Do	1 Fr
2 Mo	2 Do	2 Sa	2 Di	2 Fr	2 Sa
3 Di Tag der Dt. Einheit	3 Fr	3 So 1. Advent	3 Mi	3 Sa	3 So
4 Mi	4 Sa	4 Mo	4 Do	4 So	4 Mo
5 Do	5 So	5 Di	5 Fr	5 Mo	6 5 Di
6 Fr	6 Mo	6 Mi	6 Sa Heilige Drei Könige	6 Di	6 Mi
7 Sa	7 Di	7 Do	7 So	7 Mi	7 Do
8 So	8 Mi	8 Fr	8 Mo	8 Do	8 Fr
9 Mo	9 Do	9 Sa	9 Di	9 Fr	9 Sa
10 Di	10 Fr	10 So	10 Mi	10 Sa	10 So
11 Mi	11 Sa	11 Mo	11 Do	11 So	11 Mo
12 Do	12 So	12 Di	12 Fr	12 Mo Rosenmontag	7 12 Di
13 Fr	13 Mo	13 Mi	13 Sa	13 Di	13 Mi
14 Sa	14 Do	14 So	14 Mo	14 Mi	14 Do
15 So	15 Mi	15 Fr	15 Mo	15 Do	15 Fr
16 Mo	16 Do	16 Sa	16 Di	16 Fr	16 Sa
17 Di	17 Fr	17 So	17 Mi	17 Sa	17 So
18 Mi	18 Sa	18 Mo	18 Do	18 So	18 Mo
19 Do	19 So	19 Di	19 Fr	19 Mo	8 19 Di
20 Fr	20 Mo	20 Mi	20 Sa	20 Di	20 Mi
21 Sa	21 Do	21 So	21 Mo	21 Mi	21 Do
22 So	22 Mi	22 Fr	22 Mo	22 Do	22 Fr
23 Mo	23 Do	23 Sa	23 Di	23 Fr	23 Sa
24 Di	24 Fr	24 So Heiligabend	24 Mi	24 Sa	24 So
25 Mi	25 Sa	25 Mo 1. Weihnachtstag	25 Do	25 So	25 Mo
26 Do	26 So	26 Di 2. Weihnachtstag	26 Fr	26 Mo	9 26 Di
27 Fr	27 Mo	27 Mi	27 Sa	27 Di	27 Mi
28 Sa	28 Di	28 Do	28 So	28 Mi	28 Do
29 So Ende der Sommerzeit	29 Mi	29 Fr	29 Mo	29 Do	29 Fr Karfreitag
30 Mo	30 Do	30 Sa	30 Di		30 Sa
31 Di Reformationstag		31 So Silvester	31 Mi		31 So Beginn der Sommerzeit

Kurzes Treffen in Person

Schreibt euch eure 3-4 Lieblingsthemen in Reihenfolge auf

Vergeben es ad-hoc

Vortragsreihenfolge ist ungefähr die Themenreihenfolge

Freitag, 11 Uhr?

Diskussiongruppe?

Telegram, WhatsApp, ...