

# TA4AIStartups

## Machbarkeitsstudie für Technikfolgen-Abschätzung in Österreichischen KI-Startups

Tanja Sinozic-Martinez, Annika Schönauer, Lisa Smith  
FORBA-Fachgespräch 2/21  
Mittwoch, 26. Mai, 16:30-17:30 (online)

## Projektfokus und Projektteam

- Ziel: Entwicklung eines Prototyps zur Technikfolgen-Abschätzung in österreichischen KI-Startups.
- Team: Annika Schönauer (FORBA), Lisa Smith und Georgia Mammes (Prewave GmbH), Tanja Sinozic-Martinez, Titus Udrea und Steffen Bettin (ITA, ÖAW)
- Förderung: FFG Ideen Lab 4.0

## Vier Hauptaspekte des Forschungsdesigns

- Analyse bestehender Methoden, Ansätze und Indikatoren für die Definition von technischen Folgen und Risiken
- Anforderungsanalyse in der KI-Start-up Community
- Literaturübersicht über KI-Ethik und soziale Auswirkungen
- Co-creation, co-development und feedback bei der Entwicklung des Tools

## Theoretischer und methodischer Hintergrund

- Antizipation, Reflexivität, Inklusion und Reaktionsfähigkeit
- Soziale, ethische und politische Dimensionen im Zusammenhang mit KI-Technologien
- AI Governance
- Integration von Verantwortung und Technolgieedimensionen in Organisationskultur und Management

## Antizipation, Reflexivität, Inklusion und Reaktionsfähigkeit

- „(...) taking care of the future through collective stewardship of science and innovation in the present“ (Stilgoe et al., 2013)
- „what if...?“ (Ravetz, 1997)
- Constructive Technology Assessment (CTA) (Rip et al., 1995), Real-Time Technology Assessment (Guston and Sarewitz, 2002)
- Beschreiben von Werten, die Frames zugrunde liegen (insb. Popper, 1963), Verantwortung außerhalb der Wissenschaft übernehmen (Wynne, 2011)
- Erweiterung der Beteiligung (Callon et al., 2009)
- Übersetzung in Handlungen, die die Gesellschaft betreffen (Mazzucato, 2016), value-sensitive design (Friedman, 1996)

## Soziale, ethische und politische Dimensionen im Zusammenhang mit KI-Technologien

- Algorithmische Entscheidungsfindung
  - Verstärkende Diskriminierung (EOP, 2016)
  - Werte implizit und den Nutzern unbekannt (Wiener, 1988), Effekte unklar (Sunstein, 2019)
- Datenerfassung, Datenschutz, Bürgerrechte
  - Profilierung, Klassifizierung (Floridi, 2021)
- KI und autonome Systeme
  - Beispielloser Ersatz menschlicher Arbeit (Frey und Osborne, 2013); Verlust von Autonomie, Menschen als einfache Maschinen (Frischmann und Selinger, 2018)
- Private Plattformen
  - Meinungsppluralität kompromittieren (Sunstein, 2009), Propaganda und Unwahrheiten verbreiten (Tufekci, 2018).

## AI Governance

- EU-Bemühungen um einen koordinierten Ansatz für KI-Innovationen (code of ethics, plan on AI) (Europäisches Parlament, 2018; EK, 2018; 2019)
- Weißbuch der Kommission zu KI (EK, 2020)
- Ähnliche Entwicklungen sichtbar z.B. in USA, UK, Österreich, Deutschland, Australien und Kanada.
- Brancheninitiativen: Labelling, Ethics codes (AI Ethics Impact Group, 2020)
- Abwägung der Stärkung der nationalen Industrien für die internationale Wettbewerbsfähigkeit mit Fragen der Sicherheit und Verantwortung.

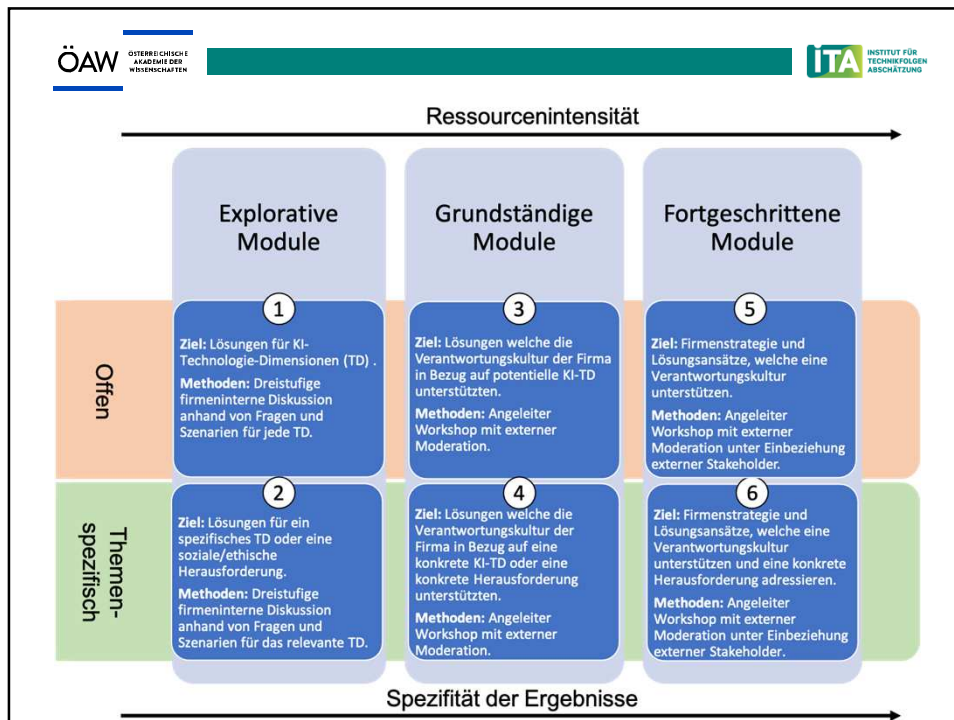
## Integration von Verantwortung und Technologiedimensionen in Organisationskultur und Management

- Im Umfeld bestehender (unverbindlicher) Initiativen und in Erwartung weiterer Regulierungen müssen Unternehmen wissen, was zu tun ist
- Herausforderungen bei der Operationalisierung
- Start-up-Landschaft wächst schnell
- Scaling-up, Disruption, Transformationspotenzial
  
- Wie lassen sich soziale und ethische Aspekte in die bestehenden Innovationsprozesse von Start-ups integrieren?



## Die modulare Toolbox für Verantwortung bei Innovationen in KI-Start-ups

- Co-Creation zwischen einem KI-Plattform-Start-up und ForscherInnen
- Schlüsselaspekte der modularen Toolbox:
  - Flexibilität
  - Spezifität der Ziele und Ergebnisse
  - Ressourceneffizienz
  - Prozessorientierung



## Szenarien (Beispiel Truth, Misinformation and Propaganda)

- Szenario 1: Gesammelte Nutzerdaten geraten unter den Einfluss von Akteuren, die an der Verbreitung von Fehlinformationen interessiert sind.
- Szenario 2: Die generierten Daten werden für schändliche Zwecke verwendet (z. B. Fake News etc.).
- Szenario 3: Die Technologie (Plattform, Dienst) wird für (politische) Fehlinformationen genutzt.