

## **Ethische, rechtliche und soziale Aspekte (ELSA) der Gestaltung von KI-Systemen und Integration der verschiedenen Akteure in den Prozess**

Sascha ALPERS<sup>1</sup>, Bettina-Johanna KRINGS<sup>2</sup>

<sup>1</sup> *FZI Forschungszentrum Informatik, Forschungsbereich Software Engineering,  
Abteilung Prozess- und Service-Orientierte Technologie  
Haid- und Neu-Straße 10-14, D-76133 Karlsruhe*

<sup>2</sup> *Institut für Technikfolgenabschätzung und Systemanalyse (ITAS),  
Karlsruher Institut für Technologie (KIT)  
Karlsstr. 11, D-76133 Karlsruhe*

**Kurzfassung:** Wesentliche Erfolgsfaktoren für die rechtlich angemessene und menschengerechte Nutzung von Künstlicher Intelligenz (KI) basieren auf der breiten Reflexion rechtlicher, ethischer und sozialer Aspekte (ELSA) im Rahmen organisatorischen Prozesse in Unternehmen und Institutionen. Ausgewählte Aspekte werden hierbei seit einigen Jahren in interdisziplinären und transdisziplinären Forschungsprojekten intensiv diskutiert und angewandt. Der folgende Beitrag ist im Rahmen des KARL-Projekt entstanden und erörtert relevante Bezugspunkte im Rahmen der Implementierung von KI in zukünftige Arbeitsumgebungen. Vor diesem Hintergrund wird die Frage bearbeitet, welche Herausforderungen Informatiker:innen im Gestaltungsprozess selbst adressieren können und zu welchen Fragen weitere Akteur:innen (Rechtswissenschaftler:innen, Ethik-Expert:innen sowie Expert:innen der sozialen Technikgestaltung) einbezogen werden sollten. Insgesamt plädieren die Autor:innen für einen prozessorientierten Ansatz, in dem unterschiedliche Aspekte der Technikgestaltung systematisch berücksichtigt werden.

### **Schlüsselwörter:**

Künstliche Intelligenz, KI, ELSA-Kriterien, ELSI, soziale Technikgestaltung, Partizipation, Stakeholder-Prozesse

## **1. Einleitung**

Neue wissenschaftliche Erkenntnisse und Anwendungsgebiete der Künstlichen Intelligenz (KI) sind seit einigen Jahren in rasantem Tempo dabei, Arbeitsprozesse sowie Arbeitsumgebungen zu verändern und neu zu gestalten. Beispielsweise hat die inzwischen weit verbreitete Plattformarbeit zu völlig neuen Organisationsformen von Arbeit in der Logistik, im Service und in der Produktion geführt (Nierling et al. 2020). In der Industrie wird die KI in der langen Tradition der Digitalisierung angewendet, um weitere vielsprechende Potentiale autonomer Systeme in der Produktion auszuloten (Hirsch-Kreinsen & Karacic 2019). Hier entwickeln so genannte „lernfähige Systeme“ zunehmend die Fähigkeit, „komplexe Verarbeitungsketten von Daten, automatische Objektidentifikationen und Sensorfunktionen auf verschiedensten Ebenen bis hin zur

Schaffung einer für die jeweilige Zielsetzung des Systems hinreichend genauen digitalen Repräsentation der Wirklichkeit realisieren und beherrschen zu können“ (Hirsch-Kreinsen & Karacic 2019:9). Hierbei geht es nicht mehr alleine darum, die Beschäftigten an neue technische Entwicklungen heranzuführen und in Modellen des „Learning-by-doing“ einzuweisen. Es geht in hohem Maße darum, neuartige Entwicklungs- und Lernprozesse im Hinblick auf die veränderten Mensch-Maschine-Interaktionen zu erkennen, zu fördern und für neue Arbeits- und Produktionsorganisationen nutzbar zu machen. Diese Entwicklungen weisen auf die steigende Komplexität moderner Arbeitsumgebungen, die den gesamten Arbeitskontext (Märkte, Organisationen, Arbeitsplatzgestaltungen) miteinbeziehen und ebenfalls stark verändern (Schröter 2019, Krings et al. 2021). Gleichzeitig nutzen Unternehmen die permanente Datenerfassung nicht nur als Grundlage für strategische teil- oder vollautomatisierte betriebliche Entscheidungsprozesse, sondern zunehmend mehr zur Leistungsbewertung, bzw. zur „Sanktionierung von Fehlverhalten“ (Christl 2021:6) ihrer Beschäftigten. Automatisierung von Steuerungsprozessen sowie die Leistungskontrolle scheinen dann auch die beiden großen Trends zu sein, die im Rahmen der KI (kritisch) diskutiert werden (Reimann & Tisch 2021).

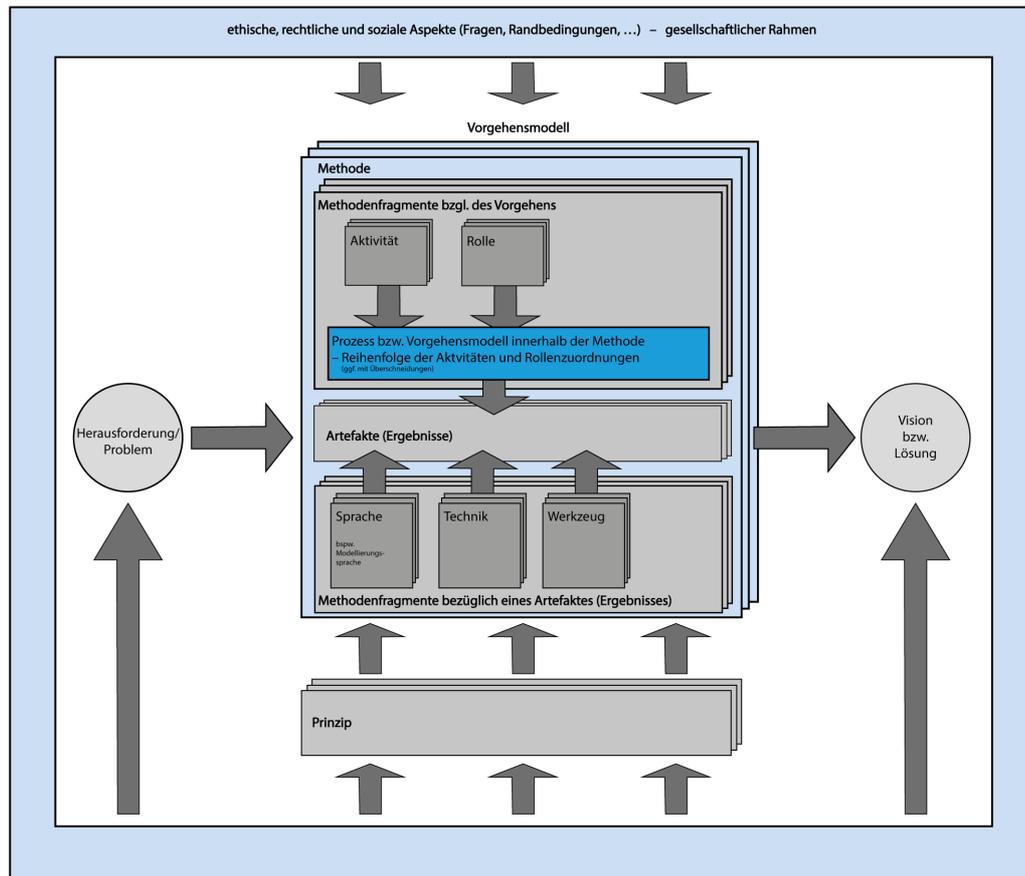
Das Verbundprojekt KARL knüpft proaktiv an diese Herausforderungen an, indem es Regeln, Konzepte und Handlungsstrategie zu KI im Praxiseinsatz erprobt (vgl. <https://kompetenzzentrum-karl.de>). Ein wichtiger Orientierungsrahmen, um diese Interaktionen zu gewährleisten, stellt das so genannte ELSI-Konzept („Ethical, Legal and Social Implications“, auch ELSA „Ethical, Legal and Social Aspects“) dar, das im Rahmen des Projektes angewendet wird. Hierbei zeigt sich jedoch, dass die menschengerechte Gestaltung von KI-Systemen eine komplexe Aufgabe darstellt, für die aktuell nur wenig Systematisierung existiert. Die weitgehend abstrakte Anforderung ethische, rechtliche und soziale Aspekte in KI-Projekten zu adressieren, wird in KARL umfassend erprobt und eine der Kernthesen lautet, dass es eine systematische Verankerung von ELSA in Vorgehensmodellen von (Informatik-) Projekten bedarf (Alpers 2022).

Vor diesem Hintergrund wird die Frage im ersten Teil (Kapitel 2) beantwortet, welche Herausforderungen Informatiker:innen im Gestaltungsprozess selbst adressieren können und zu welchen Fragen weitere Akteur:innen einbezogen werden sollten. Insgesamt plädieren die Autor:innen jedoch für einen prozessorientierten Ansatz, in dem unterschiedliche Aspekte der Technikgestaltung systematisch berücksichtigt werden. Diese Anforderungen werden im zweiten Teil des Beitrags (Kapitel 3) vorgestellt und abschließend (Kapitel 4) diskutiert.

## **2. Die Rolle der Informatiker:innen im ELSA Gestaltungsprozess der KI**

Die individuelle Verantwortung von Informatiker:innen und weiteren Akteur:innen bei der Gestaltung von Informationssystemen wird in der Regel innerhalb von Organisationen wahrgenommen. Das Zusammenwirken vieler Akteur:innen kann dazu führen, dass einzelne Akteur:innen ihre Verantwortung als gering einschätzen, weil sie ihren Einfluss auf das Ergebnis als gering wahrnehmen oder andere Akteur:innen (ggf. in anderen Rollen) als verantwortlich wahrnehmen. Bierhof und Rohmann (2017) sprechen von einer „Verantwortungsdiffusion“. Debatin (2016) spricht von einer „systematischen Verdünnung der Verantwortung“ und schlägt ein Konzept von Hans Lenk vor.

Mit dem „Modell der distributiven Mitverantwortung“ (vgl. Lenk 2015:147) werden spezifische Verantwortungen den Beteiligten zugewiesen. Alpers (2022) diskutiert, wie Vorgehensmodelle der Informatik angepasst werden müssten, damit sowohl Individuen strukturell zur Übernahme von Verantwortung befähigt und zur Einbeziehung von ELSA ermutigt werden, als auch Organisationen bzw. Verantwortliche für Vorgehensmodelle ihrer eigenen Verantwortung gerecht werden (vgl. Abbildung 1).



**Abbildung 1:** ELSA und Vorgehensmodelle, ELSA als Rahmenbedingungen die sich auf konkrete Elemente neuer bzw. angepasster Vorgehensmodelle auswirken müssen (Alpers, 2022).

Zu den in (Alpers 2022) vorgeschlagenen Konsequenzen gehören neue oder angepasste Aktivitäten sowie ein klareres Rollenverständnis. Allerdings ist es nicht immer wirtschaftlich sinnvoll oder möglich, alle Rollen mit spezifischen Expert:innen zu besetzen. Gegenwärtig fehlt es an Hilfestellungen für Organisations- und Projektverantwortliche, um zu entscheiden, für welche Projekte und Rollen die Besetzung mit spezifischen Expert:innen notwendig bzw. sinnvoll ist und in welchen Projekten vorhandene Akteur:innen ELSA hinreichend integrieren können. In einem ELSA-integrierenden Vorgehensmodell kann es daher eine frühe Prozessaktivität sein, die ELSA-Komplexität des Vorhabens und die ELSA-Kompetenz der Beteiligten daraufhin zu überprüfen, ob Rollen mit ggf. externen Expert:innen besetzt werden sollen oder ob vorhandene Akteur:innen externe Beratung benötigen. Um diese erste Entscheidung sinnvoll ohne Einbeziehung externer Expert:innen treffen zu können, muss daher eine entsprechende Grundkompetenz und ein Problembewusstsein vorhanden sein.

### 3. ELSA-Kriterien als normativer Orientierungsrahmen für KI-Anwendungen

Wie in Kapitel 2 dargestellt, scheint es von großer Bedeutung, dass Informatiker:innen, die in die Gestaltung von KI-Systemen involviert sind, ein Problembewusstsein über die vielfältigen sozialen, ethischen und politischen Folgen dieser Systeme in den jeweiligen Anwendungsfeldern entwickeln. Angesichts der Komplexität dieser Systeme wird ein normativer Orientierungsrahmen über gemeinschaftlich anerkannte Werte im Hinblick auf die Umsetzung von KI in Arbeits- und Lebenskontexten, unerlässlich (Spiekermann 2019). Dieser Rahmen ist weit gesteckt und reicht von ethischen über rechtlichen hin zu sozialen Aspekten der sozialen Wirklichkeit. Diese werden zwar in diskursiven Debatten analytisch getrennt, sind aber in sozialen Zusammenhängen zutiefst verwoben und bedingen sich gegenseitig in hohem Maße. Vor diesem Hintergrund verwundert die Vielzahl der inzwischen veröffentlichten (Ethik-) Guidelines auf nationaler und internationaler Ebene im Hinblick auf KI wenig. Obgleich alle die Intention verfolgen, einen umfassenden Orientierungsrahmen für KI-Entwicklung und -Anwendungen zur Verfügung zu stellen, unterscheiden sich die Guidelines doch erheblich sowohl in ihren thematischen Schwerpunkten, als auch auf der analytischen Ebene der Betrachtung.

Exemplarisch sei hier der Bericht „Nachhaltigkeitskriterien für Künstliche Intelligenz“ vom *Institut für ökologische Wirtschaftsforschung (IÖW)* genannt, das den thematischen Schwerpunkt auf die Frage richtet, inwiefern KI für die Umsetzung einer nachhaltigen Entwicklung genutzt werden könne (Rohde et al. 2021). Die Autor:innen erarbeiten einen Katalog mit dreizehn Kriterien sowie mit diesen verbundenen Indikatoren für eine nachhaltige KI, die sie intensiv diskutieren. Der Bericht richtet sich an Entwickler:innen der KI und zeichnet sich durch den Anspruch aus, einen ganzheitlichen Begriff von Nachhaltigkeit zu entwickeln, d.h. KI soll sowohl nachhaltig wirken – Stichwort „nachhaltige KI“ - als auch Nachhaltigkeit ermöglichen (Rohde et al.:19). Entlang der drei Nachhaltigkeitsdimensionen sozial, ökologisch und ökonomisch werden schrittweise sechs soziale, vier ökologische sowie drei ökonomische Kriterien für eine Nachhaltige KI entwickelt. Die sozialen Kriterien umfassen unter anderem „Transparenz und Verantwortung“, „Nicht-Diskriminierung und Fairness“ sowie „Selbstbestimmung und Datenschutz“. Das Kriterium „Transparenz und Verantwortungsübernahme“ wird noch ergänzt durch zwei Indikatoren, welche den Inhalt des Kriteriums weiter konkretisieren. Der Faktor Transparenz wird durch den ersten Indikator noch ergänzt um „Erklärbarkeit und Prüfbarkeit des Modells“ und der zweite lautet „Informationsmöglichkeit zur Funktionsweise des Systems“ (Rohde et al.:58). Aus diesen Indikatoren werden schließlich weitere Subindikatoren abgeleitet und formuliert. Diese Subindikatoren werden danach unterschieden, ob sie sich an die entwickelnde oder die einsetzende Organisation richten. Die Autor:innen adressieren in dieser Ethik-Guideline also gezielt bestimmte Akteur:innen, die KI weiterentwickeln und anwenden und zeigen deren Verantwortlichkeit in der Gestaltung auf (Rohe et al. 2021).

Angesichts der sich rapide vertiefenden Krisen wie der Klimakrise und einer fortschreitenden sozialen Polarisierung weltweit, weisen diese Ansätze auf ganzheitliche Problemlagen. So sollte der zukünftige technologische Wandel nicht zuletzt hinsichtlich seiner sozialen, politischen und ökologischen Implikationen im Hinblick auf diese Problemlagen befragt werden. Die konkrete Bewertung von KI erfordert vor den globalen Herausforderungen neue Ansätze, welche die Koexistenz von technologischen

und sozialen Innovationen erkennen und fördern (Hengstschläger 2020). Nur so können die vielseitigen Bedürfnisse in den Lebens- und Arbeitsrealitäten der Menschen, aber auch politische Handlungsräume auf einer gesellschaftlichen Ebene wahrgenommen, bearbeitet und umgesetzt werden.

Aber auch die Einführung von KI in spezifischen Arbeitskontexten wirft eine Reihe von ethischen, rechtlichen und sozialen Fragestellungen auf. Im Rahmen der KI als Automationstechnologie wären hier zunächst mögliche Beschäftigungseffekte ihres Einsatzes abzuschätzen: Bedroht dieser Einsatz bestehende – oder im Sinne intergenerationaler Gerechtigkeit auch zukünftige – Arbeitsplätze? Vor welchen Zielen wird die KI in die Arbeitsumgebungen angepasst. Wie wird „gute“ Arbeit im Sinne einer sinnstiftenden Arbeit auf der Basis neuer Technologien gefördert? Welches sind zukünftige Herausforderungen „guter Arbeit“ und wie können KI-Anwendungen hier sinnvolle Beiträge leisten?

Die empirische Relevanz der KI in Arbeitskontexten zeigt schon jetzt, wie sehr es durch den Einsatz von KI zu Veränderungen von Arbeitsorganisation und -inhalten kommt, die sich in vielen Branchen und Sektoren nachteilig für die Beschäftigten auswirkt (Schaupp 2021). Viele Beispiele der algorithmischer Steuerungsfunktionen weisen vehement darauf, wie die Autonomie der Beschäftigten in ihren Handlungs- und Entscheidungsräumen beschränkt werden. Die Unterwerfung von Beschäftigten unter Formen algorithmischen Managements, das sie zu bloßen Befehlsempfängern degradiert, die freie Entfaltung ihrer Befähigungen einschränkt und ständige Leistungskontrolle zumindest technisch ermöglicht, ist hierbei insbesondere zu problematisieren (Schaupp 2021, Christl 2021).

Die Gestaltung der Arbeitswelt unter Bedingungen von Technikimplementierungen findet in den Unternehmen regelgeleitet auf der Basis des Betriebsverfassungsgesetzes sowie dem Arbeits- und Gesundheitsschutz statt. Allerdings machen die Expert:innen vielerorts darauf aufmerksam, dass die Interaktion von Mensch und Maschine in Zeiten der Digitalisierung und KI neue relevante ethische, rechtliche und soziale Fragen aufwirft. Vor diesem Hintergrund sollten Entwickler von ADM-Prozessen (Algorithmic Decision Making) rechtliche und ethische Prämissen schon bei der Programmierung und beim ADM-Design berücksichtigen. Diese Anforderung wird seit Jahren im Rahmen des Konzeptes Ethics-by-Design erfolgreich diskutiert und erprobt (Rudschies et al. 2021). Darüber hinaus richtet sich der Blick auch auf den Vorgang der Implementierung von KI in Arbeitsprozessen. Nach Schröter findet hier ebenfalls ein qualitativer Wandel statt, der in hohem Maße „partizipations- und mitbestimmungsorientierte wissenschaftliche Ausarbeitungen als Unterstützung von Betriebs- und Personalräten realisiert“ (Schröter 2019:143). Diese Anforderung könne jedoch nur in prozessorientierten Ansätzen erfolgreich umgesetzt werden. Mögliche Gestaltungswege, die aktuell in Forschungsprojekte durchgeführt werden, orientieren sich einerseits an ELSA-Kriterien, andererseits treten sie in einen Prozess ein, der als „agiles kooperatives Changemanagement“ mit seiner nach vorne offenen Gestaltungshaltung“ (Schröter 2019:148) bezeichnet werden kann. Nur über beides, ein verbindlich kommunizierter Wertekanon sowie ein ergebnisoffenes, prozessorientiertes Vorgehen, kann eine „vorausschauende Technikgestaltung“ (Schröter 2019:150) gelingen.

#### 4. Diskussion

Der Beitrag zeigt mit dem Fokus ethischer Fragestellungen in KI-Projekten die Not-

wendigkeit auf, Projektverantwortlichen und -durchführenden geeignete Hilfestellungen zur Verfügung zu stellen. Gleichzeitig wird die Komplexität aufgrund der Vielschichtigkeit der Fragestellungen deutlich. Insbesondere aus der Perspektive der Gestaltung der Arbeitswelt ist neben der Einbeziehung der Betroffenen eine systematische Betrachtung der Auswirkungen der Technikgestaltung auf die zukünftige Arbeitswelt notwendig. Dabei sind neben der sozialen Technikgestaltung auch ethische und rechtliche Fragestellungen zu berücksichtigen.

## 5. Literatur

- Alpers S (2022). Notwendigkeit der Integration von ethischen, rechtlichen und sozialen Aspekten in die gängigen Vorgehensmodelle für IT-Projekte. In: Fazal-Baqaie M, Linssen O, Volland A, Yigitbas E, Engstler M, Bertram M, Kalenborn A (Hrsg.), Projektmanagement und Vorgehensmodelle 2022 - Virtuelle Zusammenarbeit und verlorene Kulturen? Bonn: Gesellschaft für Informatik, 171–179.
- Bierhoff HW, Rohmann E (2017). Diffusion von Verantwortung. In: Heidbrink L, Langbehn C, Loh J (Hg.) Handbuch Verantwortung. Springer Reference Sozialwissenschaften. Springer VS, Wiesbaden.
- Debatin B (2016). Verantwortung. *Communicatio Socialis*, 49, 68–73.
- Christl W (2021): Digitale Überwachung und Kontrolle am Arbeitsplatz. Von der Ausweitung betrieblicher Datenerfassung zum algorithmischen Management? Wien: Cracked Labs – Institut für Kritische Digitale Kultur.
- Hirsch-Kreinsen H, Karacic A (2019) Potentiale und Grenzen der Anwendung autonomer Systeme. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt. Bielefeld: Transkript.
- Hengstschläger M (2020) (Hg.) Digitaler Wandel und Ethik. Salzburg, München: Ecowin,
- Krings B-J, Moniz A B, Frey P. (2021) Technology as enabler of the automation of work? Current societal challenges for a future perspective of work. *Revista Brasileira de Sociologia*, 9 (21), 206–229.
- Lenk H (2015): Human-soziale Verantwortung. Zur Sozialphilosophie der Verantwortlichkeiten. Borchum, Freiburg: Projektverlag (Kultur & Philosophie, Band 11)
- Nierling L, Krings B-J, Küstermann L (2020) The Landscape of Crowd work in Germany. An overview of the scientific and public discourse. Karlsruher Institut für Technologie (KIT).
- Reimann M, Tisch A (2021) Arbeits- und Beschäftigungsqualität in digitalisierten Arbeitswelten. Sonderheft *Soziale Welt* 4/2021, 72 Jg., 371–593.
- Rohde F, Wagner J, Reinhard P, Petschow U, Meyer A, Voß M, Mollen A (2021). Nachhaltigkeitskriterien für künstliche Intelligenz: Entwicklung eines Kriterien- und Indikatorensets für die Nachhaltigkeitsbewertung von KI-Systemen entlang des Lebenszyklus, Schriftenreihe des IÖW 220/21, Berlin.
- Rudschies C, Schneider I, Simon J (2021). Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities. *The International Review of Information Ethics*, 29.
- Schaupp S (2021) Technopolitik von unten. Algorithmische Arbeitssteuerung und kybernetische Proletarisierung. Berlin: Matthes & Seitz.
- Schröter W (2019) Der mitbestimmte Algorithmus. Arbeitsweltliche Kriterien zur sozialen Gestaltung von Algorithmen und algorithmischen Entscheidungssystemen. In: Schröter W (Hg.) Der mitbestimmte Algorithmus. Gestaltungskompetenz für den Wandel der Arbeit. Mössingen-Talheim: talheimer, 101–150.
- Spiekermann Sarah (2019) Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert. München: Droemer

**Danksagung:** Die Publikation entstand im Projekt „Kompetenzzentrum KARL – Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe“. Das Projekt wird durch das Bundesministerium für Bildung und Forschung (BMBF) im Programm „Zukunft der Wertschöpfung – Forschung zu Produktion, Dienstleistung und Arbeit“ (Förderkennzeichen: 02L19C250) gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor:innen. Ein ganz besonderer Dank gilt den Mitgliedern des so genannten Gestaltungsfeldes 5 im KARL-Projekt, in dessen Rahmen die Inhalte des vorliegenden Artikels entwickelt und kritisch diskutiert wurden.