Unterstützung der Wissensarbeit durch Künstliche Intelligenz – Anforderungen an die Gestaltung maschinellen Lernens

Gerhard SATZGER[1], Niklas KÜHL[2], Andrea MARTIN[3]

[1,2] Karlsruhe Institute of Technology (KIT)
Kaiserstraße 89, D-76133 Karlsruhe
[3] IBM Deutschland GmbH
Hollerithstraße 1, D-81829 München

Kurzfassung: Entwicklungen der Informationstechnologie stellen zunehmend Möglichkeiten bereit, den menschlichen „Wissensarbeiter“ durch kognitive Assistenzsysteme auf Basis Künstlicher Intelligenz (KI) in seinen Entscheidungen oder Aktionen zu unterstützen. Im vorliegenden Beitrag wollen wir beleuchten, welche Anforderungen an die zugrundeliegenden maschinellen Lernverfahren zu stellen sind, um die individuelle sowie gesellschaftliche Akzeptanz solcher „Augmented Intelligence“ zu gewährleisten.

Schlüsselwörter: Künstliche Intelligenz, Wissensarbeit, Knowledge Worker, kognitive Assistenz, maschinelles Lernen, Dienstleistungswirtschaft

1. Künstliche Intelligenz als Basis kognitiver Assistenz

Mitte November 2018 ging das Bild des deutschen Astronauten Alexander Gerst um die Welt, in dem er sich auf der ISS mit einem medizinballgroßen, frei im Raum schwebenden „Assistenten“ unterhält. Dieser „Crew Interactive Mobile Companion“ (CIMON) wurde im Auftrag des DLR von Airbus entwickelt und setzt insbesondere IBM Watson als KI-Technologie ein: Ziel ist es, „mit seinen Kameras, Sensoren, Mikrofonen und Prozessoren nicht nur zu "sehen", zu "hören", zu "verstehen" und zu "sprechen", sondern auch vielfältige Informationen sowie Experiment- und Reparaturanleitungen darzustellen und zu erklären“ (DLR 2018) – mithin ein sehr populäres Beispiel für die Art von Assistenz, die mittlerweile auch für hochqualifizierte Wissensarbeiter wie die ESA-Astronauten konzipiert und umgesetzt werden kann.

Abseits effekthascherischer Science-Fiction-Filme bezeichnet Künstliche Intelligenz – bzw. der englische Begriff Artificial Intelligence (AI) - die Möglichkeiten, menschliches Denken und Tun entweder zu replizieren oder aber alternativ rational zu unterstützen (Russel & Norvig 2015). KI wendet Verfahren des maschinellen Lernens an (Kühl et al. 2019), um die kognitiven Fähigkeiten des Menschen zu erweitern, d.h. alle Prozesse, in denen „Sinneseindrücke transformiert, verdichtet, verarbeitet, gespeichert, wieder aufgefunden und benutzt werden“ (Niesser 1967, S. 4, Übersetzung durch Autor). In diesem Sinne könnte man auch von *„Augmented Intelligence“* als der Erweiterung der menschlichen Intelligenz oder von *kognitiver Assistenz* (Demirkan et al. 2015, Siddike & Kohda 2018) als der Unterstützung bei Erfassung und Verarbeitung von Informationen sprechen: Der einzelne Konsument kommt heute bereits in den Genuss dessen - ob durch Sprachassistenten wie Amazons Alexa oder Echo in einfachen Situationen des täglichen Lebens, bei der Unter-

stützung in Finanzanlagen (Jung et al. 2018) oder gar durch humanoide Roboter bei der Kaufberatung (Nestlé 2014).

Bedeutung erlangt diese insbesondere auch für Wissensarbeiter bzw. „Knowledge Worker“ (Drucker 1959), die vor allem für die Anwendung Ihres Wissens bezahlt werden: Für sie stellt sowohl die Reichhaltigkeit des zur Verfügung stehenden Wissens als auch die Geschwindigkeit auf dessen Zugriff einen wichtigen Wettbewerbsvorteil dar (Fischbach & Putzke 2012). Damit leisten kognitive Assistenzssysteme auf Basis künstlicher Intelligenz einen wesentlichen Beitrag insbesondere auch zu einer innovationsorientierten Dienstleistungswirtschaft (Ernst & Zühlke-Robinet 2018): So erhalten Kundendiensttechniker heute Unterstützung bei der Lösung von Störungen (Herterich et al. 2015, Satzger & Hottum 2015), Finanzmanager für Umsatz- oder Cash Flow-Prognosen (Satzger et al. 2015, Blanc & Setzer 2015), Banksachbearbeiter bei der Bearbeitung von Kreditanträgen (Abdou & Pointon 2011) oder Ärzte bei der Diagnose von Patienten (Miller & Brown 2017).

Andererseits machen nicht zuletzt spektakuläre Missbrauchsfälle wie z.B. derjenige der Firma Cambridge Analytica im US-Wahlkampf (Beuth & Horchert 2018) deutlich, dass der Einsatz der Technologie geeigneter Prinzipien und Rahmenbedingungen bedarf, deren Einhaltung dann auch wiederum geprüft und nachgewiesen werden muss. Im Folgenden entwickeln wir einen Überblick über die entsprechenden Anforderungen und erläutern diese im Einzelnen.

2. Anforderungen

Für eine verantwortliche Unterstützung von Wissensarbeitern sollten maschinelle Lernverfahren als Kern kognitiver, KI-basierter Assistenzsysteme einer Reihe von Anforderungen genügen. Deren Erfüllung ist sowohl auf der Ebene des einzelnen Wissensarbeiters als auch auf gesellschaftlicher Ebene bedeutsam, um Vertrauen in die Technologie und in die entsprechende Entscheidungsunterstützung zu bekommen und letztlich auch für die Akzeptanz kognitiver Assistenzsysteme zu sorgen. Abbildung 1 stellt die im folgenden angerissenen Themenkomplexe im Überblick dar:

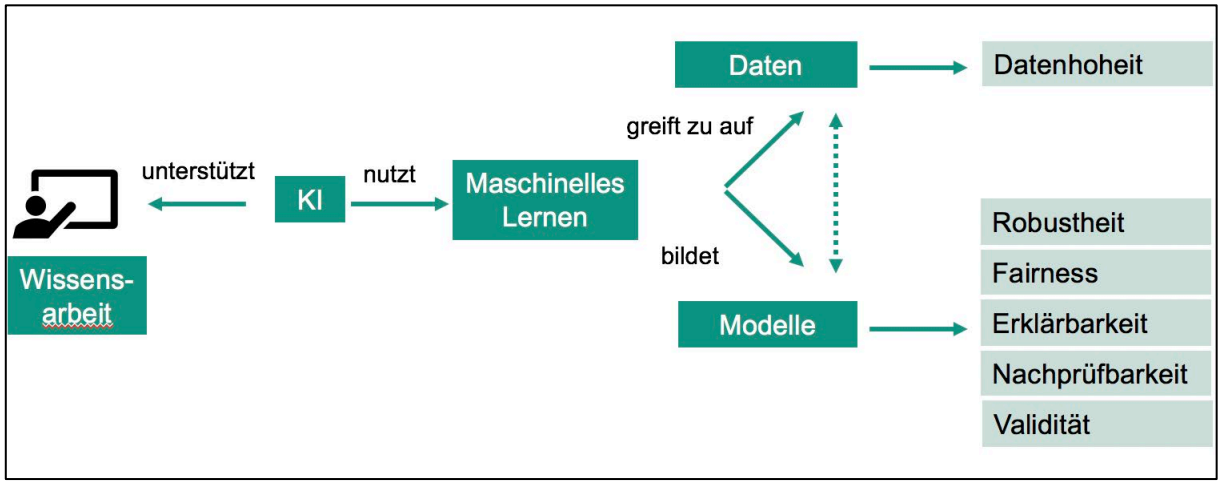

Abbildung 1: *Künstliche Intelligenz (KI) und Anforderungen an das maschinelle Lernen.*

Die den Wissensarbeiter unterstützende Künstliche Intelligenz (KI) greift auf maschinelle Lernverfahren zurück. Diese werden entweder in sog. “simple-reflex

agents“ genutzt, die – einmal trainiert – jeweils unverändert für die Assistenzerbringung angewandt werden oder alternativ in sog. „learning agents“, die im laufenden Betrieb durch entsprechende Rückkopplung laufend weiter adaptiert werden (Hirt et al. 2019). Beiden Ansätzen gemein ist, dass auf Basis existierender Datensets Modelle gebildet werden, die dann auf weitere Sachverhalte in Form neuer Daten angewandt werden. Sowohl in Hinblick auf die zum „Training“ des Modells genutzten Daten als auch auf die Bildung des Modells können Anforderungen für eine verantwortungsvolle Nutzung zur Unterstützung von Wissensarbeit erhoben werden.

2.1 Datenhoheit

Die Einsichten künstlicher Intelligenz beruhen auf Daten – in jüngster Zeit haben gerade die Verbreitung von Sensorik (z.B. im Rahmen des Internet of Things oder speziell von „Industrie 4.0“) oder von sozialen Netzwerken maßgeblich dazu beigetragen, die Menge verfügbarer, auswertbarer Daten explosionsartig ansteigen zu lassen. Eine Studie der IDC schätzt die 2020 verfügbare Datenmenge auf 44 Zettabyte bzw. 44 Trillionen Gigabyte (IDC 2014): Speicherte man diese Menge auf handelsüblichen 2TB-Festplatten und reihte man diese aneinander, so ergäbe dies eine Kette mit einer Länge, die der siebenfachen Entfernung der Erde zum Mond entspricht. Dabei sind neben strukturierten auch zunehmend unstrukturierte Daten Gegenstand der Auswertung – wie z.B. im gerade im Dezember 2018 gestarteten Pilotprojekt der Mannheimer Polizei zur verhaltensanalysierenden Videoüberwachung öffentlicher Brennpunkte (Stadt Mannheim 2018). Ein verantwortlicher Umgang mit Assistenzsystemen sollte voraussetzen, dass die entsprechende Unterstützung nur auf Daten beruht, die auch zu diesem Zweck freigegeben oder öffentlich verfügbar sind. Anbieter kognitiver Assistenzsysteme haben sich bereits explizit dazu verpflichtet, wie z.B. die IBM Corporation in ihren „Trust and Transparency Principles“ (IBM 2018). Dies bedingt zum einen, dass Schutzmechanismen gegen Datenmissbrauch (wie im obigen Fall von Cambridge Analytica der unrechtmäßige Zugriff auf Facebook-Nutzerdaten) gestärkt werden müssen, zum anderen, dass die Transparenz für die Verwendung von Daten gestärkt wird Diese wurde z.B. im Mai 2018 durch das Inkrafttreten der Datenschutz-Grundverordnung (DSGVO) bereits deutlich gefördert; laufende Entwicklungen versuchen, diese auch permanent nachweisbar und zertifizierbar zu machen (BMWi 2018). Darüber hinaus wird an maschinellen Lernverfahren gearbeitet, die die Erarbeitung von Modellen ermöglichen, *ohne* die zugrundeliegenden Daten preisgeben zu müssen, z.B. über sog. „Meta-Learning“ (Hirt & Kühl 2018).

2.2 Robustheit

Der Wissensarbeiter muss sich auf die Werthaltigkeit der ihm zur Verfügung gestellten Unterstützung verlassen können. Dies impliziert nicht nur, dass das Assistenzsystem insgesamt den sonst üblichen Sicherheitsanforderungen, die an IT-Systeme zu stellen sind, genügen muss. Insbesondere ein System mit einem „learning backend“ (Kühl et al. 2019), das also kontinuierliches Lernen zulässt, ist so auszulegen, dass es sich robust gegen Eingriffe von außen selbst dann verhält, wenn die Manipulationsversuche auf zulässigem Wege, also über die Lernkomponente des Systems erfolgen. So wurde 2015 Microsofts Kommunikationsbot „Tay“, der aus der Kommunikation mit Twitter-Nutzern lernen sollte, von einer Gruppe gezielt mit Hass-

kommentaren gefüttert, so dass er von Microsoft bereits nach wenigen Stunden aus dem Netz genommen werden musste (Sickert 2016).

2.3 Fairness

Das zugrundeliegende Modell darf keine Verzerrungen aufweisen, die z.B. aufgrund unausgewogener Trainingsdaten entstehen können: Nehmen wir an, ein Unternehmen versucht, aus den Charakteristika bisherig eingestellter Kandidat/inn/en auf deren Erfolg zu schließen, um so dem „Recruiter“ als Wissensarbeiter Hilfe bei der Evaluation von Bewerbern zu geben. Hat das Unternehmen z.B. bislang vornehmlich Betriebswirtschaftsabsolvent/inn/en eingestellt, ist die zur Verfügung stehende Datenbasis unausgewogen, und das Modell mag vornehmlich wieder solche empfehlen („Status-Quo-Bias“). Für Ingenieurkandidat/inn/en mögen keine Erfahrungen vorliegen und daher auch keinerlei positive Aussage getroffen werden können – obwohl diese ggf. besser geeignet sein könnten. Aus ethischer Sicht sind diese Verzerrungen umso kritischer, je geschützter eine betroffene Personengruppe (z.B. aufgrund von Hautfarbe, Herkunft, Geschlecht, etc.) ist. Insbesondere muss das Modell dafür Sorge tragen, dass das betreffende Merkmal eben *keinen* Beitrag zur einer Bewertung oder Empfehlung liefern darf. Hierzu gibt es eine Reihe von Metriken, die bei der Erstellung oder Anwendung eines Modells laufend überprüft werden können – für die Entdeckung von Diskriminierung z.B. den sog. Disparate Impact, der die Abweichung geschützter und nicht-geschützter Gruppen analysiert (Feldman et al. 2015). Gleichzeitig geben entsprechende Werkzeuge (z.B. IBM’s AI OpenScale) auch Hinweise, wie solche Verzerrungen eliminiert werden können.

2.4 Erklärbarkeit

Die Verständlichkeit und (zumindest partielle) Nachvollziehbarkeit der vom Assistenzsystem angebotenen Unterstützung ist ein wichtiges Akzeptanz- und vertrauensbildendes Kriterium. Die Nutzung interpretierbarer Algorithmen (z.B. Support Vector Machines im Gegensatz zu neuronalen Netzen), die Visualisierung und die Bereitstellung kontextspezifischer Erklärungen können hierzu wesentlich beitragen. Wird z.B. eine ärztliche Diagnose durch ein Assistenzsystem vorgeschlagen, so sollte der Arzt aufgrund der Symptomatik den Befund nachvollziehen können – selbst dann, wenn er z.B. bei seltenen Erkrankungen diesen nicht selbst – ohne Assistenzsystem – so hätte erheben können.

2.5 Nachprüfbarkeit

Ebenso sollten die Schritte zur Entwicklung und Anpassung eines Modells dokumentiert und ggf. auch nach festgelegten Kriterien geprüft werden. Hier geht es nicht um die Nachprüfbarkeit eines einzelnen Ergebnisses (wie in 2.4), sondern um den Ablauf der Modellentwicklung als solche. Diese sog. „lineage“ könnte z.B. auch formal über „Suppliers‘ Declarations of Conformity“ (Hind et al. 2018) oder durch unabhängige Stellen zertifiziert werden. Insbesondere muss dokumentiert werden, wer mit welchen Daten zu welchem Zeitpunkt und mit welchen Verfahren Modelle erstellt oder geändert hat – z.B. um unausgewogene Modelle, die also einen „Bias“ aufweisen, identifizieren oder die Aktualität der Basisdaten überprüfen zu können.

2.6 Validität

Schließlich gilt es, die Gültigkeit der jeweils erstellten Modelle auch im Zeitablauf zu gewährleisten. Ändern sich die zugrundeliegenden Verhältnisse, die sog. „ground truth“, muss das Modell neu trainiert bzw. kalibriert werden. Dazu ist es notwendig, zu erkennen, wann solche Veränderungen eingetreten sind und – in der einschlägigen Terminologie – ein „concept drift“ vorliegt (Baier 2019). Dies ist insbesondere dann kritisch, wenn die Assistenz in einer Empfehlung für eine Aktion mündet, die selbst wieder die Datenlage beeinflusst: so wird z.B. im Bereich der industriellen Instandhaltung prädiktive Wartung betrieben, die (hoffentlich) dann auch das Ausfallverhalten der Maschinen und damit die Grundlage der Modellbildung beeinflusst.

3. Diskussion: „Mensch-mit-Maschine“?

Insgesamt lässt sich damit konstatieren, dass Künstliche Intelligenz durchaus mächtige (und sehr wettbewerbsrelevante) Assistenzsysteme für den Wissensarbeiter hervorbringen wird. Insofern sollten die entsprechenden Technologien auch im Fokus sowohl der Dienstleistungswissenschaft als auch des Innovationsmanagements in der Praxis stehen. Für eine verantwortliche Unterstützung des Wissensarbeiters unter pragmatischen wie auch ethischen Prinzipien ist die Erfüllung der im Vorangegangenen diskutierten Anforderungen wesentlich.

Forschungsseitig erwachsen daraus eine ganze Reihe von Aufgaben: so gilt es, Verfahren zu entwickeln, um die einzelnen Anforderungen jeweils zu prüfen und zu belegen (wie z.B. oben erwähnt über die Funktionalitäten entsprechender Software). Interessant wäre es darüber hinaus, dies einerseits von unabhängiger Stelle und andererseits kontinuierlich vornehmen zu können (Lins et al. 2016). Des Weiteren werden Erkenntnisse benötigt, um den richtigen Grad der Symbiose zwischen dem menschlichen Wissensarbeiter und der technischen Assistenz für den jeweiligen Anwendungsfall festzulegen – und die geeigneten Schnittstellen zu schaffen.

Mithin gilt es, im internationalen Wettbewerb das Potential kognitiver Assistenzsysteme bzw. der KI-unterstützen Wissensarbeit zu erschließen – und dies in verantwortlicher Weise zu tun. Die Diskussion sollte sich also nicht – wie häufig im Kontext von KI - um „Mensch oder Maschine“ drehen, sondern um die geeignete verantwortliche Gestaltung von „Mensch-mit-Maschine“.

4. Literatur

Abdou HA, Pointon J (2011) Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature. Intelligent Systems in Accounting, Finance and Management 18(2-3):59-88

Baier L, Kühl N, Satzger G (2019) How to Cope with Change? Preserving Validity of Predictive Services over Time. Proceedings of the Hawaii International Conference on System Sciences

Beuth P, Horchert J (2018) Was treibt eigentlich Cambridge Analytica, http://www.spiegel.de/netzwelt/netzpolitik/cambridge-analytica-das-steckt-hinter-der-datenanalyse-firma-a-1198962.html (aufgerufen am 8.12.18)

Blanc SM, Setzer T (2015) Analytical debiasing of corporate cash flow forecasts. European Journal of Operational Research 243(3):1004–1015.

BMWi (2018) AUDITOR - European Cloud Service Data Protection Certification, https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/Einzelprojekte/einzelprojekte_auditor.html (abgerufen am 9.8.18)

Demirkan H, Bess C, Spohrer J, Rayes A, Allen D, Moghaddam Y (2015) Innovations with Smart Service Systems: Analytics, Big Data, Cognitive Assistance, and the Internet of Everything. Communications of the Association for Information Systems 37(1):733-752

DLR (2018) CIMON - ein "Cyberkollege" für Alexander Gerst, https://www.dlr.de/dlr/desktopdefault.aspx/tabid-10081/151_read-28678/#/gallery/31256 (aufgerufen am 6.12.18)

Drucker PF (1959) The Landmarks of Tomorrow. Harper and Row: New York

Ernst G, Zühlke-Robinet K (2018) Dienstleistungen – Wissenschaft und Forschung, Arbeit und Innovation, nomos:Baden-Baden

Feldman N, Friedler SA, Moeller J, Scheidegger C, Venkatasubramanian S (2015) Certifying and Removing Disparate Impact. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 259-268

Fischbach K, Putzke J (2012) Wissensarbeiter, http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Wissensmanagement/Wissensorganisation--Instrumente-der-/Wissensarbeiter (aufgerufen am 9.12.18)

Herterich M, Peters C, Übernickel F, Brenner W, Neff A (2015) Mobile Work Support for Field Service: A Literature Review and Directions for Future Research. Wirtschaftsinformatik Proceedings, 10.

Hind M, Mehta S, Mojsilovic A, Nair R, Ramamurthy KN, Olteanu A, Varshney KR (2018) Increasing Trust in AI Services through Supplier's Declarations of Conformity, arXiv: 1808.07261

Hirt R, Kühl N (2018) Cognition in the Era of Smart Service Systems: Interorganizational Analytics through Meta and Transfer Learning. Proceedings of the International Conference on Information Systems

IBM (2018) IBM's Principles for Trust and Transparency, https://www.ibm.com/blogs/policy/trust-principles/ (aufgerufen am 8.12.18)

IDC (2014) The Digital Universe of Opportunities, Rich Data and the Increasing Value of the Internet of Things, https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm (aufge-rufen am 8.12.18)

Jung D, Dorner V, Weinhardt C, Pusmaz H (2018) Designing a robo-advisor for risk-averse, low-budget consumers. Electronic Markets 28(3):367-380

Kühl N, Goutier, Hirt R, Satzger G (2019) Machine Learning in Artificial Intelligence: Towards a Common Understanding. Proceedings Hawaii International Conference on Systems Sciences (in Druck)

Lins S, Grochol P, Schneider S, Sunyaev A (2016), Dynamic Certification of Cloud Services: Trust, but Verify!, IEEE Security & Privacy, 14(2):66-71

Miller DD, Brown EW (2017) Artificial Intelligence in Medical Practice: The Question to the Answer? American Journal of Medicine 131(2):129-133

Neisser U (1967) Cognitive Psychology. Appleton-Century-Crofts:New York

Nestlè (2014) Nestlé to use humanoid robot to sell Nescafé in Japan, http://www.nestle.com/media/news/nestle-humanoid-robot-nescafe-japan (abgerufen 6.12.18).

Russel SJ, Norvig P (2015) Artificial Intelligence: A Modern Approach, 3. Aufl. Pearson:Boston

Satzger G, Holtmann C, Peter S (2015) Advanced Analytics im Controlling – Potenzial und Anwendung für Umsatz- und Kostenprognosen. Controlling 27(4/5):229–235.

Satzger G, Hottum P (2015) Management von Interaktionsqualität in industriellen Dienstleistungsnetzwerken – Ein Service-Analytics–Ansatz für die Störungsbearbeitung, Zeitschrift für betriebswirtschaftliche Forschung 67(Suppl69):150-173.

Sickert T (2016) Vom Hipster-Mädchen zum Hitlerbot, http://www.spiegel.de/netzwelt/web/microsoft-twitter-bot-tay-vom-hipstermaedchen-zum-hitlerbot-a-1084038.html (aufgerufen am 6.12.18)

Siddike MAK, Kohda Y (2018) Co-creating Value in People's Interactions with Cognitive Assistants: A Service-System View. Journal of Creating Value 4(2):255-272

Stadt Mannheim (2018) Startschuss für die intelligente Videoüberwachung, https://www.mannheim.de/de/nachrichten/startschuss-fuer-intelligente-videoueberwachung (aufgerufen am 8.12.18)

Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit interdisziplinär
analysieren – bewerten – gestalten

65. Kongress der
Gesellschaft für Arbeitswissenschaft

Professur Arbeitswissenschaft
Institut für Technische Logistik und Arbeitssysteme
Technische Universität Dresden

Institut für Arbeit und Gesundheit
Deutsche Gesetzliche Unfallversicherung

27. Februar – 1. März 2019

GfA-Press

Bericht zum 65. Arbeitswissenschaftlichen Kongress vom 27. Februar – 1. März 2019

Professur Arbeitswissenschaft, Institut für Technische Logistik und Arbeitssysteme, Technische Universität Dresden;
Institut für Arbeit und Gesundheit, Deutsche Gesetzliche Unfallversicherung, Dresden

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2019
ISBN 978-3-936804-25-6

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Screen design und Umsetzung
© 2019 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de