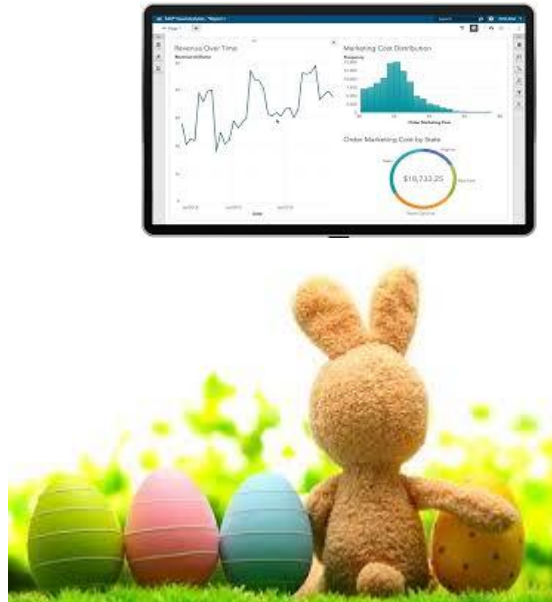


Business Intelligence und Data Mining

Dr. Robert Stahlbock
Universität Hamburg
Institut für Wirtschaftsinformatik (IWI)

- **Organisation**
- Thematische Einordnung
Business Intelligence und Analytische Informationssysteme





- Institut für Wirtschaftsinformatik (Prof. Dr. Stefan Voß)
- Tel.: 42838-3063
- VMP5, Raum 1039 (B/C, 1. Stock)
- E-Mail: robert.stahlbock@uni-hamburg.de
- <https://www.bwl.uni-hamburg.de/iwi/team/mitglieder/robert-stahlbock.html>
- Sprechstunde: wegen Corona derzeit nur digital (E-Mail; Telefon o.ä. nach Absprache)

- zweijährige Banklehre nach Abitur und Grundwehrdienst
- Studium Betriebswirtschaftslehre (Uni Hamburg) mit Schwerpunkten Betriebswirtschaftliche Datenverarbeitung, Informatik und Unternehmensforschung (Operations Research) → Dipl.-Kfm.
- studienbegleitend: Mitarbeit in Bank und Unternehmensberatung
- Forschung und Lehre als Wissenschaftlicher Mitarbeiter und Lehrkraft für besondere Aufgaben am Institut für Wirtschaftsinformatik der Uni Hamburg, → Promotion Dr. rer. pol.
- Forschung/Lehre am Institut für Wirtschaftsinformatik/Uni Hamburg (insbes. Business Intelligence, Data Science, OR, Container-Logistik)
- General Chair International Conference on Data Science (ICDATA), Las Vegas, USA
<https://icdata.org>; 2020: The 16th Conference...
- Mitgründer u. Senior Consultant BI³Slab
- freiberuflich Dozent an privater Fachhochschule seit 2003/04
Kostenrechnung, Controlling, Grundlagen EDV, Standardsoftware MS Office, Supply Chain Management, Transport und Logistik, insbes. Beschaffungslogistik, Datenstrukturen und Algorithmen, Informationsmanagement, E-Business, IT-Management

Business Intelligence und Data Mining

- Master; Zielgruppe: WiInf, BWL, ...
- Lehrkonzept: überwiegend Vorlesung
- Vorlesung: Erarbeitung verschiedener Themengebiete, teilweise Zeigen von Software (evtl.)
- Zusätzlich: Literaturstudium (z.B. bestimmte Fachartikel)

- **NEUE HERAUSFORDERUNG: Corona...**
- → b.a.W. ausschließlich digitale Lehre
 - asynchrone Screencasts u.ä.
(synchron bei 260 TN nicht praktikabel)
 - für Fragen: E-Mail

■ Folien

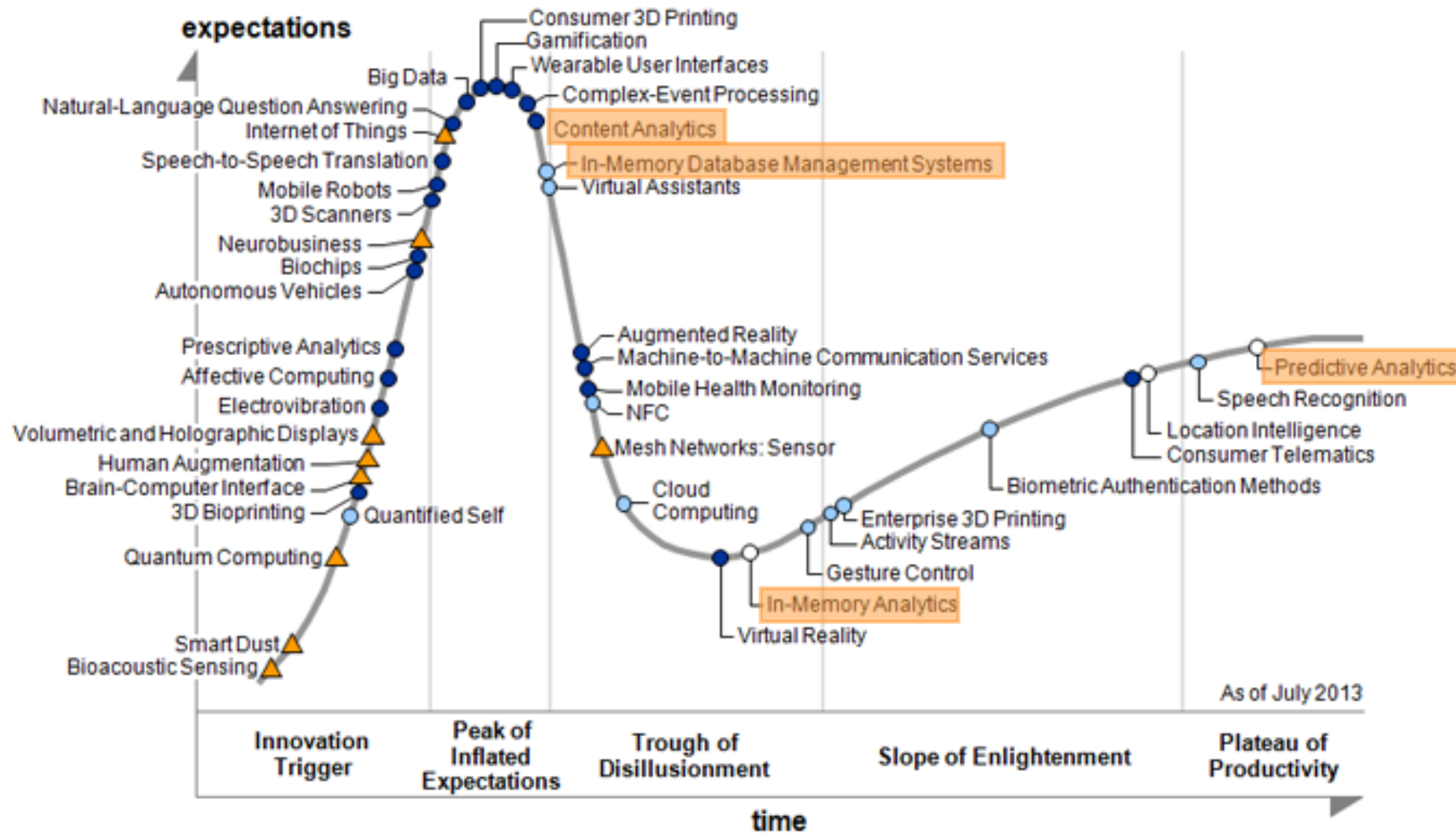
- ❑ Veranschaulichung wichtiger Konzepte und Hintergrundinformationen;
- ❑ **Überblick** über Themengebiete
- ❑ Zeigen von Dingen, die man tun kann oder sollte, aber auch die man nicht tun sollte
- ❑ Bereitstellung über Stine

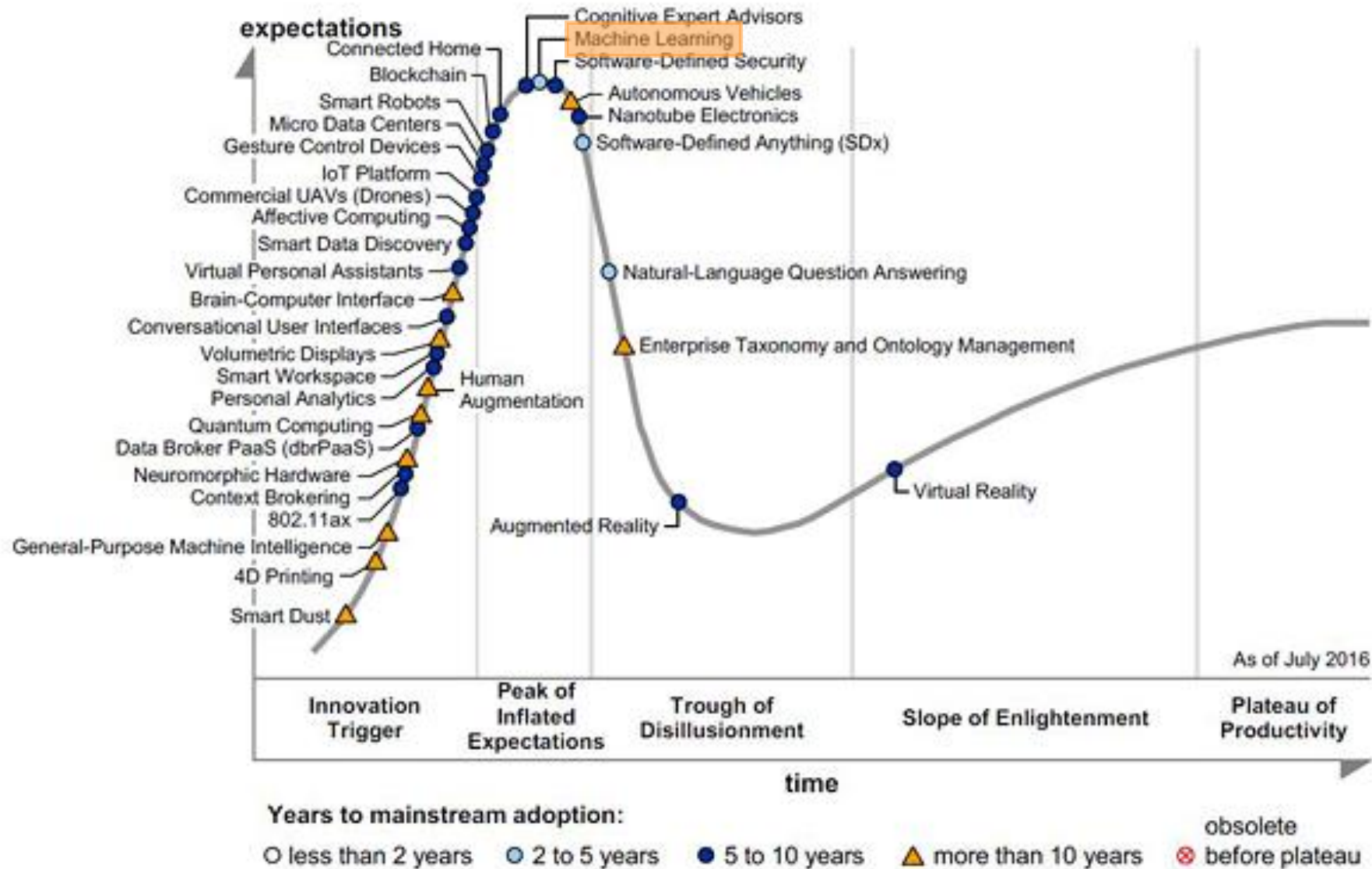
■ Prüfung

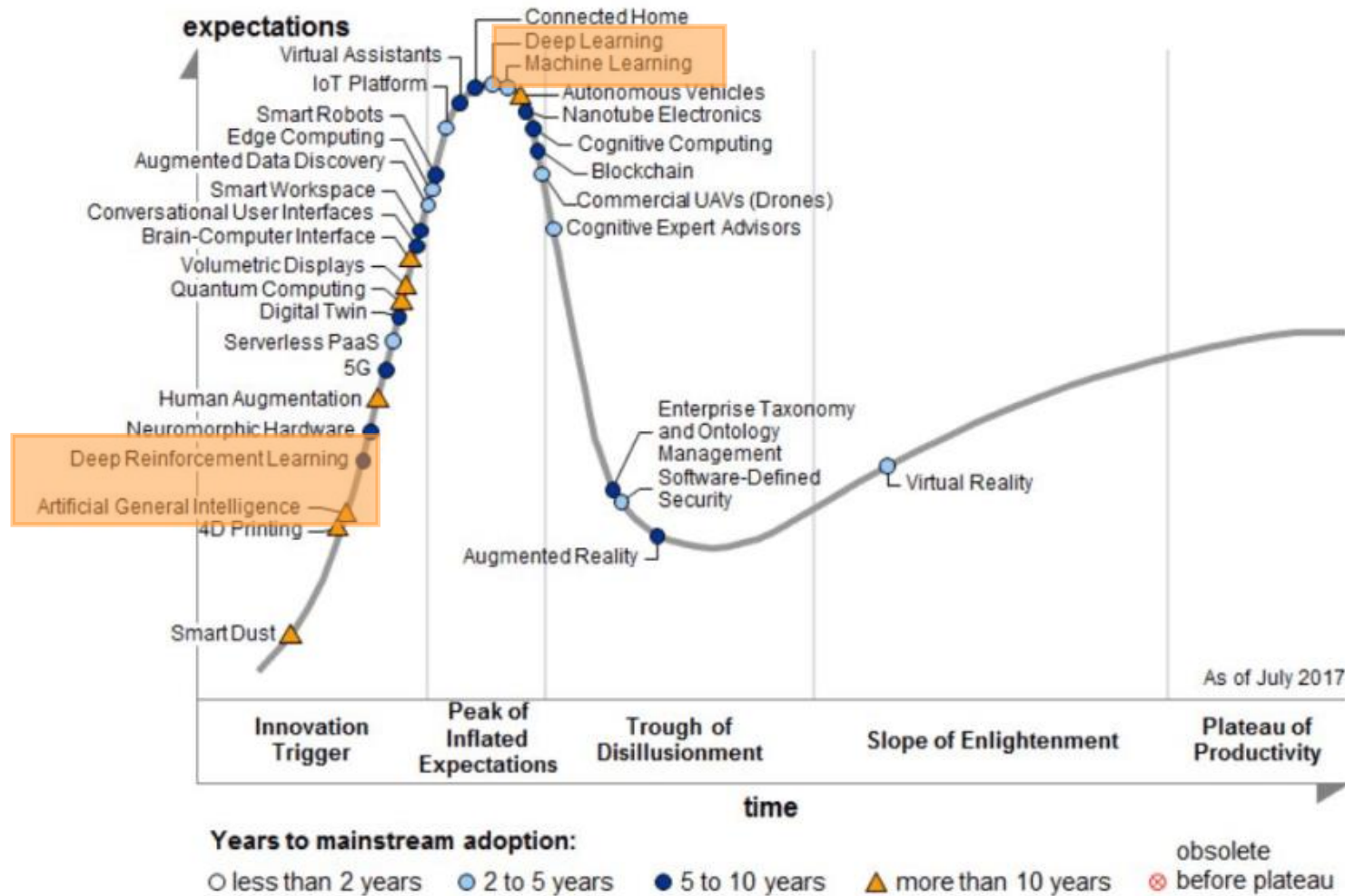
- ❑ normalerweise: Abschlussklausur (90 min), Credits: 6
Details zur Klausurdurchführung können noch nicht genannt werden
- ❑ relevant:
 - Vorlesungen
 - Literaturstudium gem. Vorgabe
... und auch sonst schaden Lesen und Eigeninitiative beim Lernen nicht 😊

- Organisation
- **Thematische Einordnung**
Business Intelligence und Analytische Informationssysteme



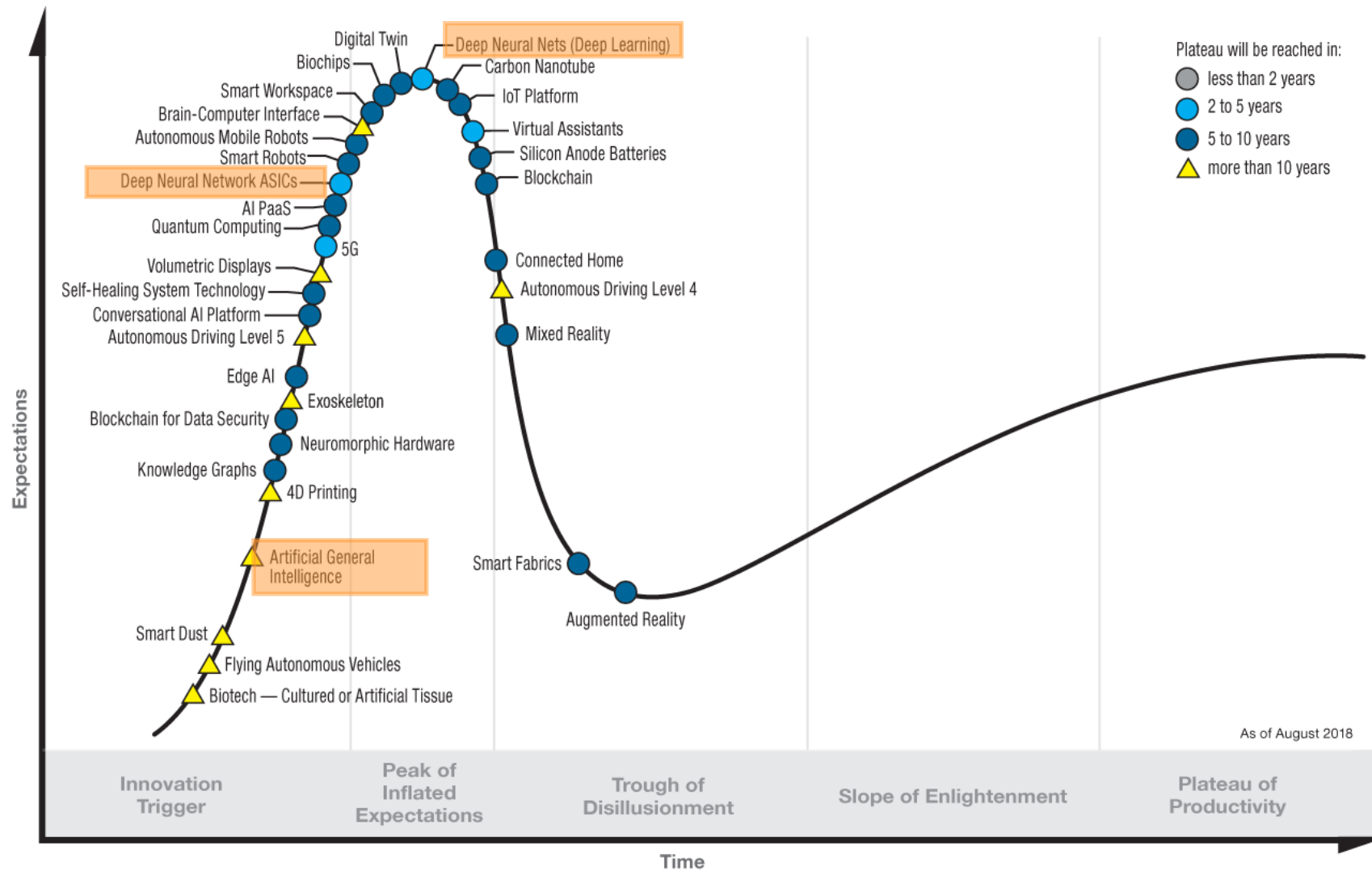






Note: PaaS = platform as a service; UAVs = unmanned aerial vehicles

Source: Gartner (July 2017)



■ Business Intelligence

- ❑ Gegenstand, Notwendigkeit, Ziele
- ❑ Analyse-, Planungs- und Entscheidungsprozesse
- ❑ Nutzergruppen und Anforderungen
- ❑ Basistechnologien und Architekturbausteine
- ❑ Analytische IS
 - Data Warehouse
 - Online-Analytical-Processing
- ❑ Wirtschaftlichkeitsüberlegungen

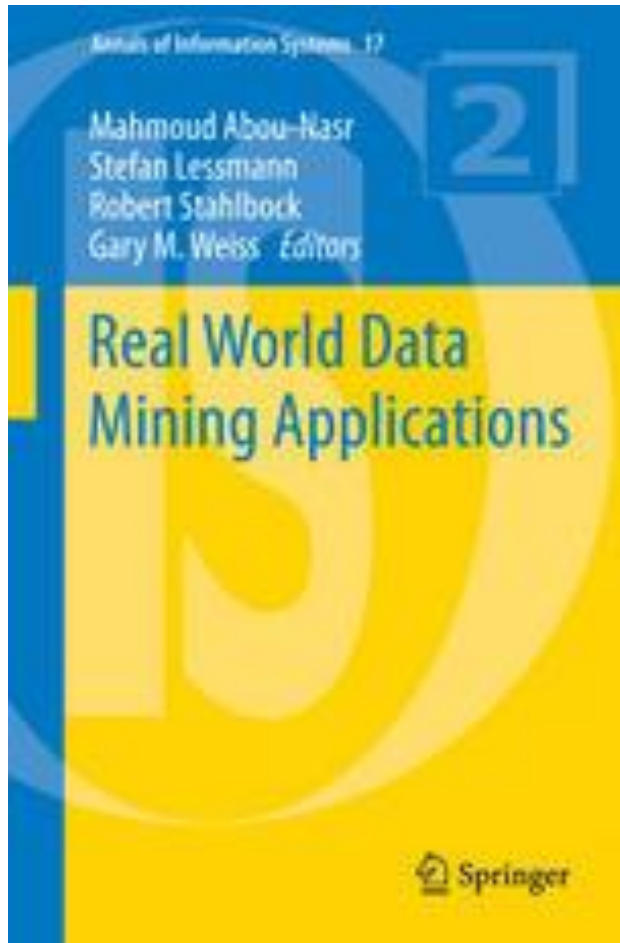
■ Data Mining

- ❑ Grundlagen, Motivation und Prozess
- ❑ Klassifikation (**A**nwendungen, **M**ethoden, **B**ewertung)
- ❑ Datenvorverarbeitung
- ❑ Segmentierende und cluster-bildende Verfahren (AMB)
- ❑ Warenkorbanalysen und Assoziationsregeln (AMB)
- ❑ Web-Mining, Text-Mining, Meinungs-/Stimmungs-Mining
- ❑ evtl. Zeitreihenanalyse

Aktuelle Entwicklungen und Trends



- P. Gluchowski, R. Gabriel, C. Dittmar. Management Support Systeme und Business Intelligence
- P. Chamoni , P. Gluchowski (Hrsg.). Analytische Informationssysteme
- M.J. Berry, G.S. Linoff. Data Mining Techniques
- H. Hippner, M. Meyer, U.L. Küsters, K.D. Wilde. Handbuch Data Mining im Marketing
- J. Han, M. Kamber. Data Mining
- T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning
- A. J. Izenman. Modern Multivariate Statistical Techniques
- Außerdem: Fachartikel, Konferenzbeiträge



Annals of Information Systems

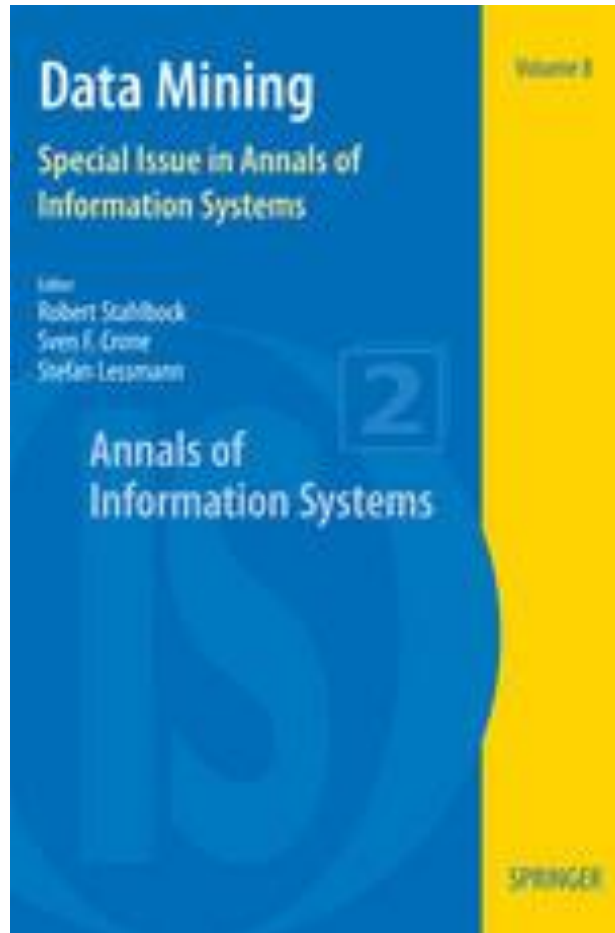
Volume 17 2015

Real World Data Mining Applications

Editors:

Mahmoud Abou-Nasr,
Stefan Lessmann,
Robert Stahlbock,
Gary M. Weiss

ISBN: 978-3-319-07811-3 (Print) 978-3-319-07812-0 (Online)



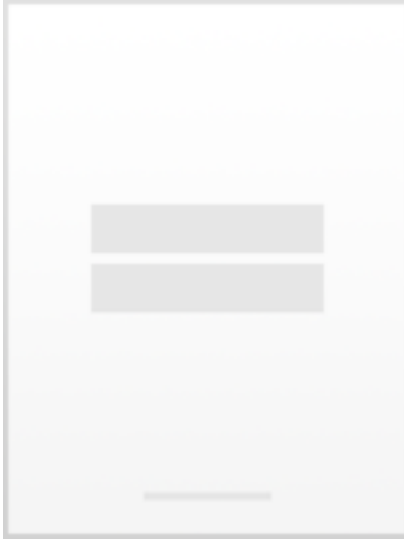
Annals of Information Systems
Volume 8 2010

**Data Mining
Special Issue in Annals of Information
Systems**

Editors:

Robert Stahlbock,
Sven F. Crone,
Stefan Lessmann

ISBN: 978-1-4419-1279-4 (Print) 978-1-4419-1280-0 (Online)



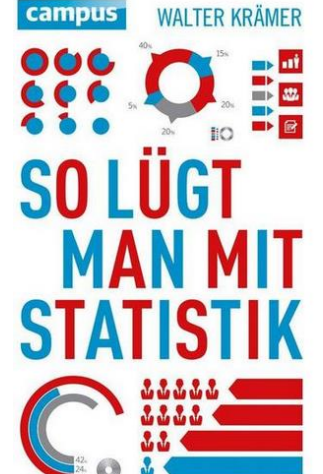
Principles of Data Science

Hrsg.: **Arabnia, H.R., Daimi, K., Stahlbock, R., Soviany, C., Heilig, L., Brüssau, K.** (Eds.)

Einführung in verschiedene Techniken, Methoden und Algorithmen

Springer (US): Transactions on Computational Science and Computational Intelligence

Juli 2020



... und auch sonstige Quellen im Web (Foren, Blogs, Social Networks, Plattformen...)

kdnuggets, ...

In der Wissenschaft ist Theorie eher fragil, Phänomenologie robust und Heuristiken und praktische Tricks sind antifragil. 😊

■ Arbeiten

□ **Masterarbeiten**

□ Verschiedene Schwerpunktsetzungen

- Fallstudie in der Praxis
- Empirische Aufgabenstellungen
- Entwicklung analytischer Software

■ Projekte/Studien

- insb. Wirtschaftsinformatiker/innen
- Gruppen von 2-4 Personen, oder auch Einzelarbeit; nach Absprache
- Praktische Aufgabenstellung, Vorträge, Abschlussarbeit
- Themenschwerpunkte wie oben

- Jährlicher Data Mining Wettbewerb für Studenten
- prudsys AG
 - ➔ <http://www.data-mining-cup.de>
- Internationales Teilnehmerfeld
- wegen Corona verschoben
- Wozu ?
 - Ruhm und Ehre
 - Preisgeld (Stand 2018):
 - 1. Preis: € 2.000
 - 2. Preis: € 1.000
 - 3. Preis: € 500



- Prognose von Retouren
- Großer Kostenfaktor im E-Commerce
 - Durchschnittliche Retourenquote: 50%
 - Neue EU-Verbraucherrichtlinie ab Juni 2014
 - Niedrige Retourenquote strategischer Wettbewerbsvorteil
- Datengrundlage
 - Historische Transaktionen
 - Kauf-/Versendezeitpunkt
 - Produktattribute (Kategorie, Preis, etc.)
 - Kundenattribute (demographische und transaktionale Daten)
 - Binäre Indikatorvariable
 - Ware wurde retourniert
 - Ware wurde nicht retourniert
- Ziel: Prognosemodell erstellen
 - Vorhersage ob Ware retourniert werden wird
 - Gegeben Kunden-/Produktmerkmale

2016: detailliertere
Weiterführung

- Prognose von Umsätzen einer Online-Apotheke mit dynamischem Pricing
- Datengrundlage
 - Historische Transaktionen (3 Monate)
 - Anschauen
 - Warenkorb
 - Kauf
 - Artikelattribute, Preis Wettbewerber u.ä.
- Ziel: Prognosemodell erstellen für die Tage des Folgemonats

- Prognose von Umsätzen einer Online-Apotheke mit dynamischem Pricing
- Datengrundlage
 - Historische Transaktionen (3 Monate)
 - Anschauen
 - Warenkorb
 - Kauf
 - Artikelattribute, Preis Wettbewerber u.ä.
- Ziel: Prognosemodell erstellen für die Tage des Folgemonats

■ Betrugserkennung bei Selbstbedienungskassen

→ „Klassifikation“

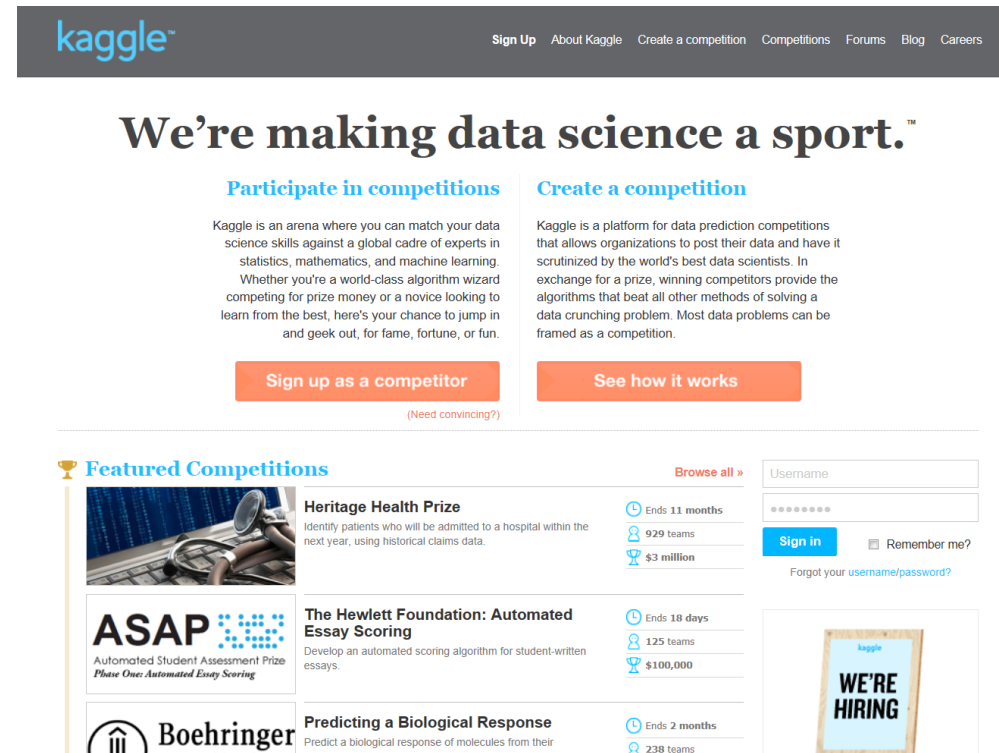
■ Datengrundlage

- ❑ Historische Transaktionen (~1.900)
- ❑ In Klassen ungleich verteilte Daten
- ❑ Asymmetrische Zielfunktion

Vorhersage	Tatsächlicher Wert		
		0 (kein Betrug)	1 (Betrug)
0 (kein Betrug)		0.00€	-5.00€
1 (Betrug)		-25.00€	5.00€

- ❑ Ziel: Prognosemodell erstellen für Betrug/Nicht-Betrug

- Prognosewettbewerb als Dienstleistung
- Analytische Aufgabenstellungen aus verschiedenen Anwendungsfeldern
 - Gesundheitswesen, Kreditwesen, Finanzprognosen, Sportprognosen, Soziale Netzwerke, Biologie, Physik, u.v.a.m.
- Attraktive Preise
(und starker Wettbewerb)
- Potenzielle Themen für Abschlussarbeiten bzw. Projekte



The screenshot shows the Kaggle homepage with a dark header containing the logo and navigation links: Sign Up, About Kaggle, Create a competition, Competitions, Forums, Blog, Careers. The main headline reads "We're making data science a sport.™". Below this are two columns: "Participate in competitions" and "Create a competition", each with a description and a "Sign up as a competitor" button. The bottom section features "Featured Competitions" with three entries: "Heritage Health Prize" (ends in 11 months, 929 teams, \$3 million), "ASAP" (Automated Student Assessment Prize, ends in 18 days, 125 teams, \$100,000), and "Predicting a Biological Response" (ends in 2 months, 238 teams). A login section on the right includes a "Sign in" button and a "Remember me" checkbox. A "WE'RE HIRING" sign is also visible.

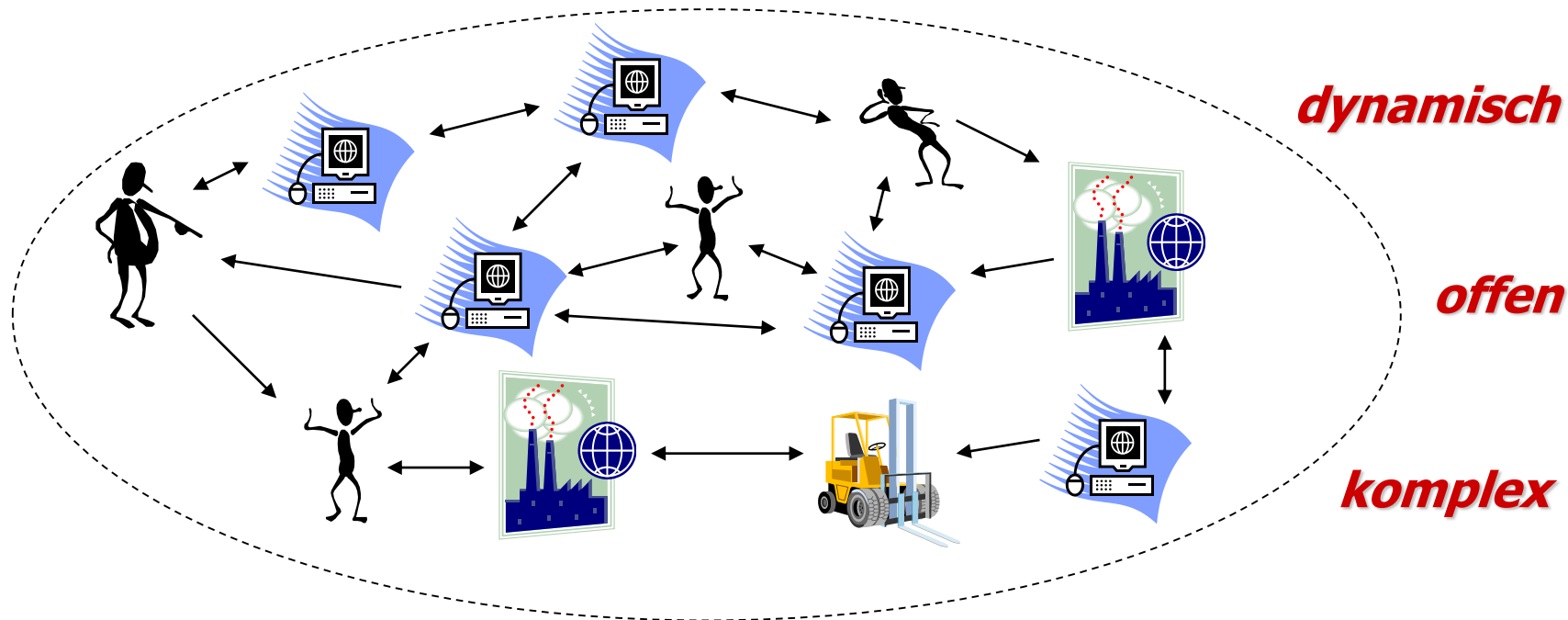
Definitionen...

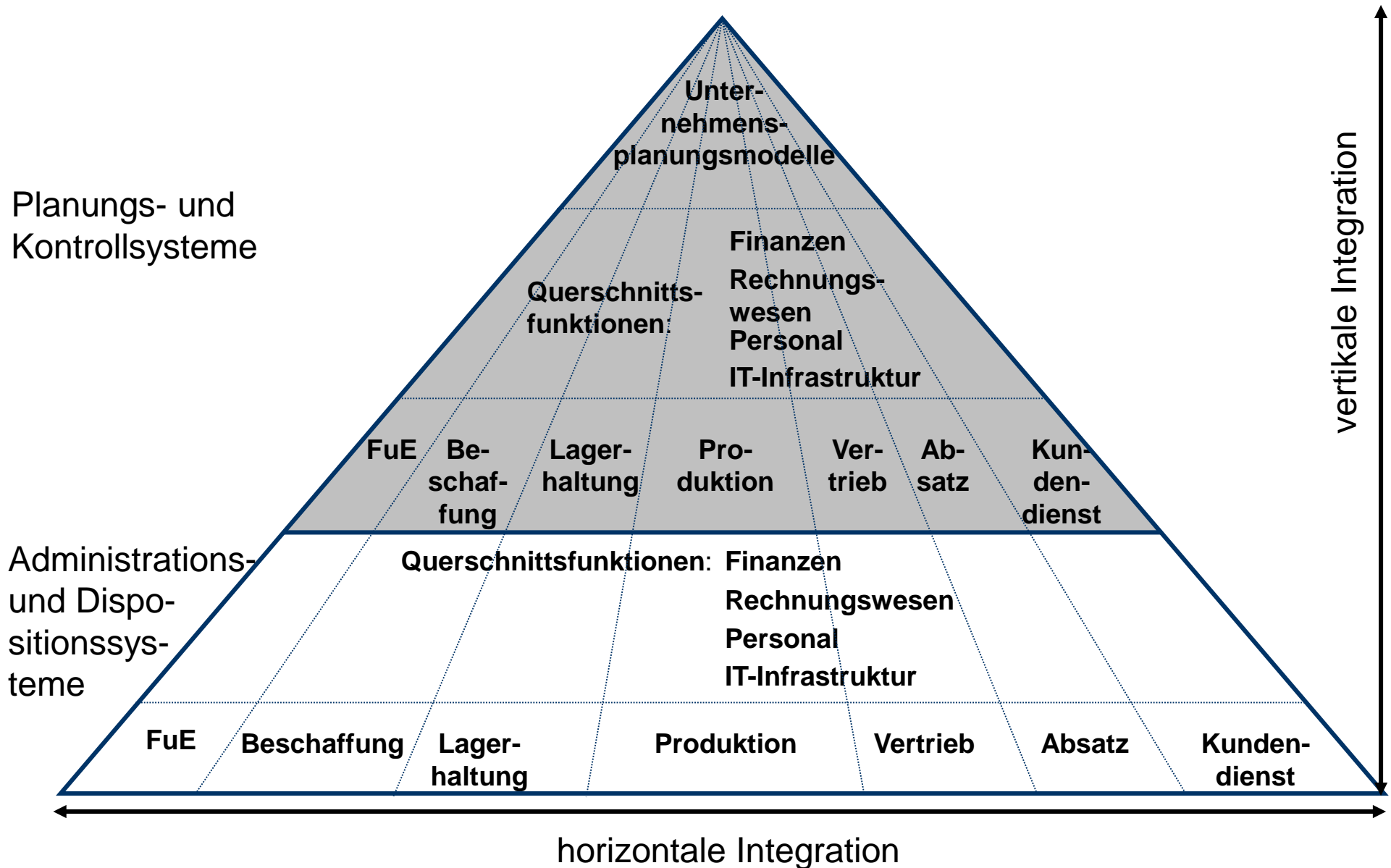
- WKWI (Wissenschaftliche Kommission Wirtschaftsinformatik, 1994)
„Gegenstand der Wirtschaftsinformatik sind Informations- und Kommunikationssysteme in Wirtschaft und Verwaltung“
- Mertens et al. (2000)
„Wirtschaftsinformatik befasst sich mit Konzeption, Entwicklung, Einführung, Wartung und Nutzung von Systemen der computergestützten Informationsverarbeitung im Betrieb.“
- Scheer (1997)
„Wirtschaftsinformatik ist die Wissenschaft von Entwurf, Entwicklung und Einsatz computergestützter betriebswirtschaftlicher Informationssysteme.“

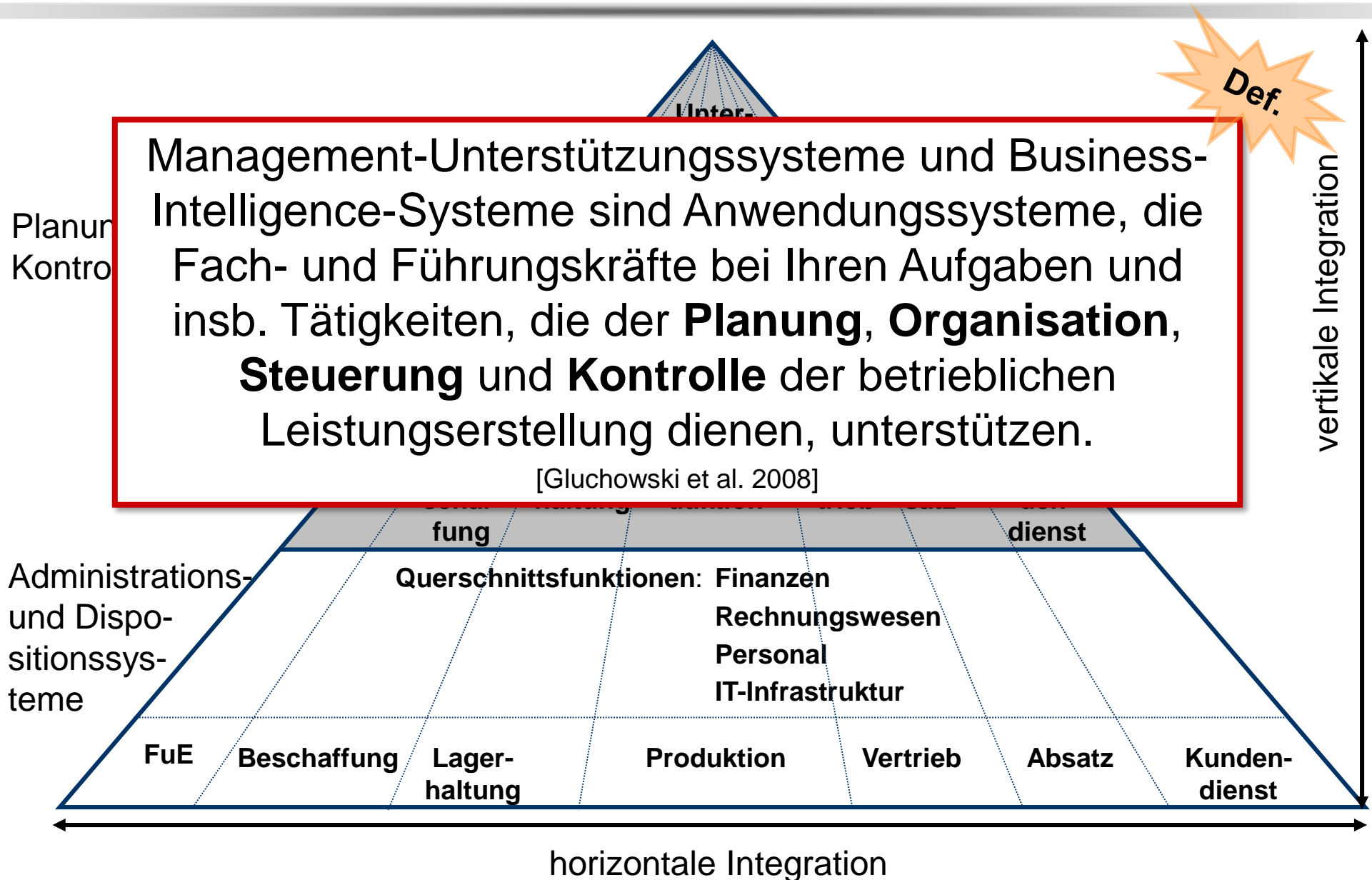


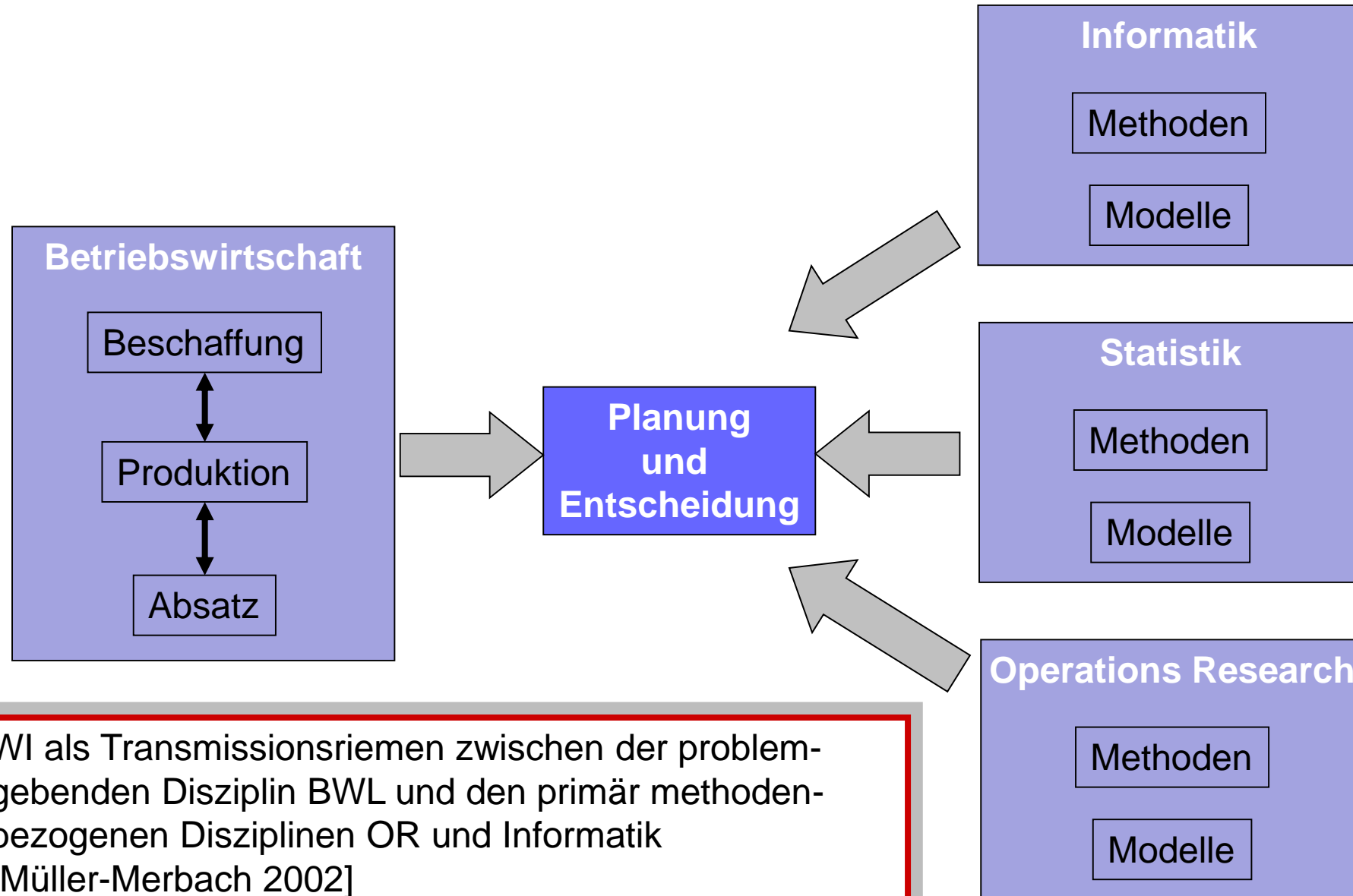
WKWI

„Informationssysteme sind soziotechnische Systeme, die menschliche und maschinelle Komponenten als Aufgabenträger umfassen, die voneinander abhängig sind, ineinandergreifen und/oder zusammenwirken.“









WI als Transmissionsriemen zwischen der problemgebenden Disziplin BWL und den primär methodenbezogenen Disziplinen OR und Informatik
[Müller-Merbach 2002]

■ Gegenstand

Unterstützung von Planung und Entscheidung
(im unternehmerischen Kontext)

■ Ansatz

- Daten-/Informationsbereitstellung
- Formale Modelle und Methoden
- Zugriff/Bereitstellung durch geeignete Anwendungssoftware

■ Aktualität/Relevanz

■ Aktualität und Relevanz

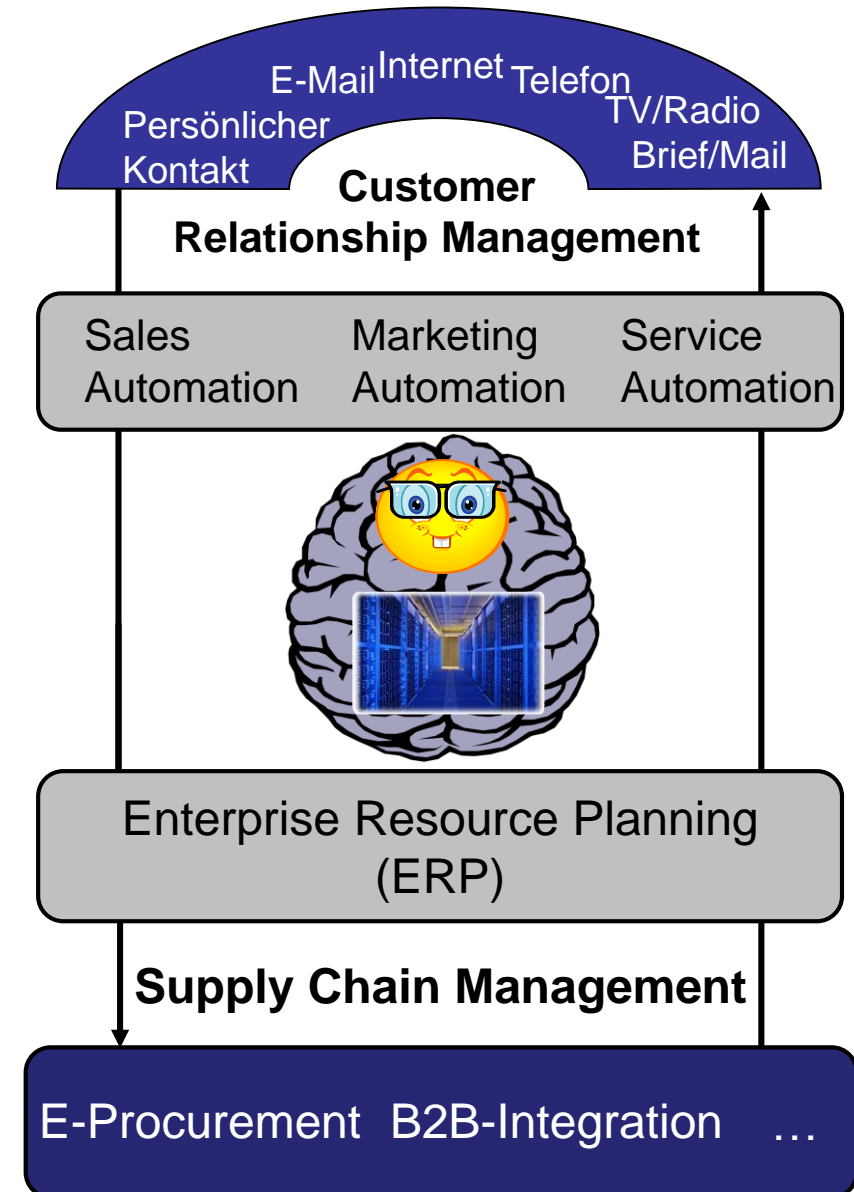
□ Externe Faktoren

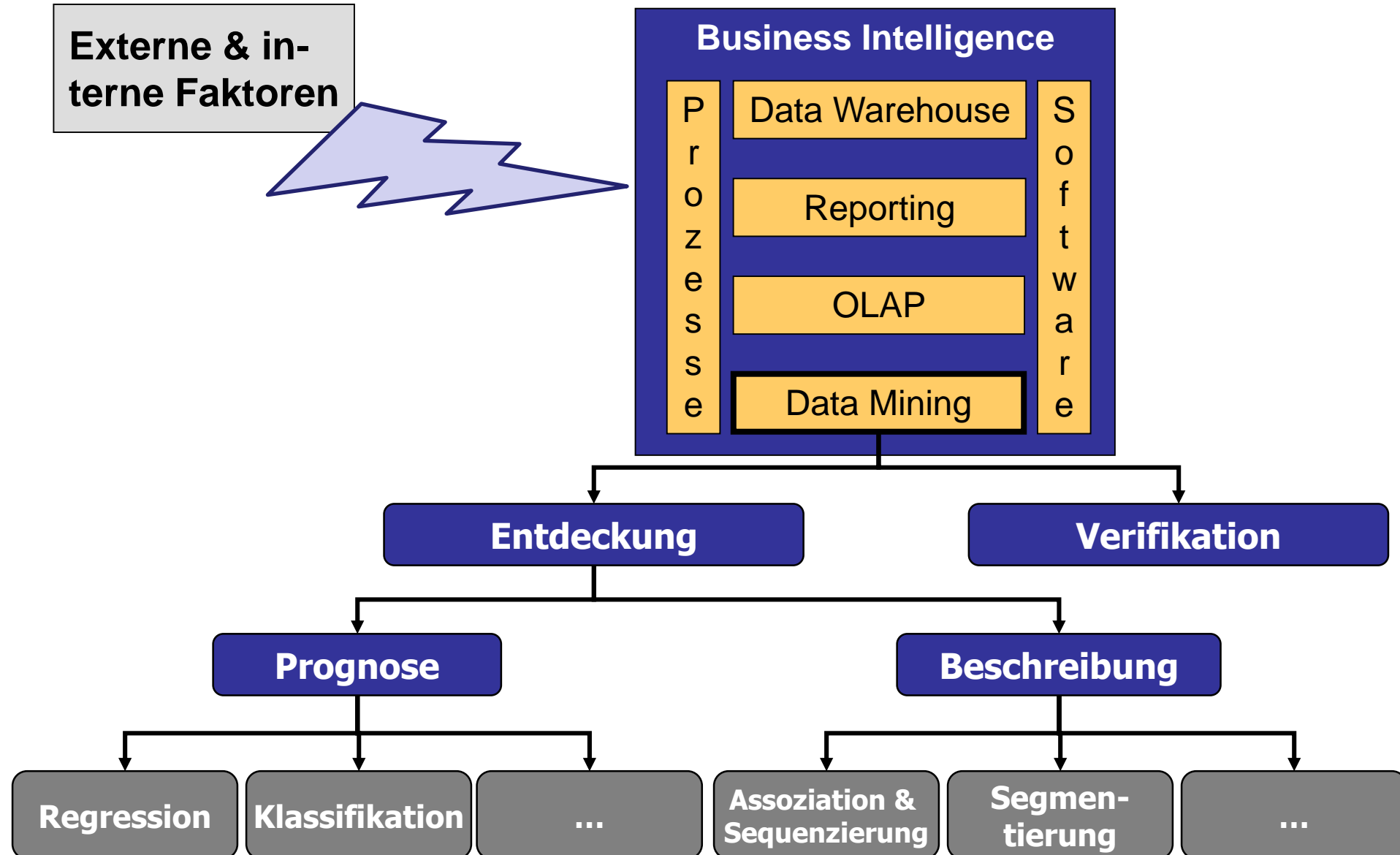
- Technologie/Internet
 - Wegfall von Handelsbarrieren
 - Homogenisierung von Produkten
- ➔ verschärfter Wettbewerb

□ Interne Faktoren

- Einsatz von Informationssystemen in allen Funktionsbereichen
- Große Datenbestände, Big Data
- Verfügbarkeit von Hard-/Software

Möglichkeit und Notwendigkeit, große Datenmengen geeignet zu speichern, zu analysieren und geschäftsrelevantes Wissen abzuleiten!





■ Business Intelligence und Data Mining

- Organisation
- Thematische Einordnung

