



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR MULTIMEDIALE
UND INTERAKTIVE SYSTEME

English
Logo

Director: Prof Dr. rer. nat. Michael Herczeg

**Explainable Artificial Intelligence:
Designing human-centric assessment system interfaces to
increase explainability and trustworthiness of artificial
intelligence in medical applications**

Master's Thesis

as part of the study program

Media Informatics

of the University of Lübeck

submitted by:

Philipp Dominik Bzdok

Issued and supervised by:

Univ.-Prof. Dr. rer. nat. Thomas Franke, Dipl.-Psych.

with support from:

Tim Schrills, M.Sc.

Lübeck, 6th October, 2021

Vorbemerkung

(Version: 2021-05-12-IMIS LaTeX Version by Philipp Bzdok)

Anbei ein kommentiertes Template für Projekt- und Abschlussarbeiten in der Medieninformatik. Die grünen Kommentare in jedem Fall vor einer (Zwischen-) Abgabe entfernen (alle!).

Alle Angaben sind eigene Einschätzung (bzw. aus den zitierten Quellen, die aber auch wieder nach eigenen Überlegungen von mir ausgewählt wurden) und entsprechend ohne Gewähr. Nordstrom's Policy gilt auch hier: "Use your own best judgment at all times."

Es ist ihre Arbeit. Sie investieren 6 Monate Ihrer Lebenszeit in die Arbeit und Sie müssen Sie vor anderen verteidigen. Sehen Sie Hinweise von anderen als Ratschläge: Ernst nehmen, aber immer für sich selbst entscheiden, was passt, was nicht und was auf ein Problem hinweist, was man dann aber anders löst. Das letzte Wort (v.a. über formelle Aspekte) hat allerdings immer der jeweilige Betreuer.

Zunächst kommen ein paar allgemeine Vorbemerkungen, die für die ganze Arbeit relevant sind (Allgemeine Punkte der Arbeit). Danach die Gliederung einer normalen Arbeit mit Anmerkungen in den entsprechenden Kapiteln bzw. Abschnitten.

Viel Erfolg beim Schreiben.

Daniel Wessel

Rundumschlag was wissenschaftliches Arbeiten betrifft:

https://www.youtube.com/watch?v=__ID7q7pXzyc

Allgemeine Punkte der Arbeit

- **Änderbarkeit von Struktur und Inhalten:** Je nach konkretem Thema kann eine andere Struktur sinnvoll sein. Dies ist ins-besondere bei der Reihenfolge der Analyse-Abschnitte der Fall, kann aber auch ganze Kapitel betreffen. Das ist insbesondere bei Masterarbeiten der Fall, die sich je nach Ausrichtung (z. B. theoretische Arbeit) stark von der sonst üblichen Form unterscheiden. Diese Struktur so früh wie möglich mit dem Betreuer klären, ihm dann per eMail zuschicken und sich kurz bestätigen lassen (generell hat man viel zu tun und erinnert sich nicht unbedingt an alle Absprachen, deswegen einfach im Anschluss kurz eine Zusam-

menfassung des Gesprächs per eMail schicken).

- **Exposé (Bachelor-/Masterarbeit) bzw. Pflichtenheft (Bachelor-/Masterprojekt)** setzt die **inhaltliche Bewertungsgrundlage**: Darin haben Sie (Arbeit) oder Ihr Betreuer (Projekt) festgelegt, was zu erreichen ist. Entsprechend genau überlegen, was man verspricht bzw. zu was man sich verpflichtet.
- **Zielgruppe**: Schreiben Sie die Arbeit so, dass andere Personen verstehen, was Sie machen — auch wenn sie keine (Medien-)Informatik studiert haben. Sprich: Beginnen Sie breit um den Einstieg zu erleichtern und formulieren Sie es allgemein verständlich (z. B. das die öffentliche Verwaltung zunehmend digitalisiert wird) und fassen Sie am Ende eines Kapitels die Punkte allgemeinverständlich wieder zusammen. Dazwischen können Sie auf ein Detaillevel runtergehen und eine Komplexität nutzen, die ein Laie nicht mehr versteht. Da der Text in Absätzen aufgebaut ist, kann der Laie (oder nicht interessierte) diese Absätze überspringen. Der Laie sollte aber zumindest im Prinzip verstehen, was Sie gemacht haben.
- **Roter Faden**: Die Arbeit logisch aufeinander aufbauen: In der Einleitung legen Sie dar, was Sie erreichen wollen und zeigen dabei auch subtil, warum das wichtig / interessant / relevant ist (*"Why should I care?"*). Das geschieht über Belege und Argumente, wann immer Sie "wichtig", "interessant" oder "relevant" im Text verwenden sind Sie dabei gescheitert. **Alle** Kapitel nach der Einleitung zeigen **logisch aufeinander folgend**, wie der Zweck der Arbeit erreicht wird.
- **Zwischenstufen-Denken**: So schreiben, dass am Ende jedes Kapitels (Einleitung, Analyse, Konzeption, Realisierung, Dialogbeispiele, Evaluation) der Leser stoppen könnte und die Arbeit selbst fortführen könnte (z. B. nach der Einleitung sich für eine anderes Analyseverfahren entscheiden, oder auf Basis der Analyse eine andere Lösung konzipieren).
- **Eigenes Fazit am Ende jeden Kapitels — Was bedeuten die Ergebnisse für die Arbeit bzw. deren Ziel?** Z. B. am Ende der Analyse kurz zusammenfassen, was die wesentlichen Punkte für die weitere Entwicklung sind — mit Fokus auf das nächste Kapitel (hier: Konzeption). Am Ende der Konzeption kurz zusammenfassen, was die wesentlichen Punkte für die weitere Entwicklung (jetzt: Realisierung) sind, etc. pp. Ein gutes Fazit ist viel Arbeit und setzt ein gut geschriebenes Kapitel voraus.

- **Professionelle Zwischenabgaben:** Wenn Sie den Text an den Betreuer geben, gehen Sie vorher kritisch drüber. Wenn Textmarken falsch gesetzt sind, die Formatierung zusammengebrochen ist, etc. dann muss sich der Betreuer da erst mal durchwühlen um zum Inhalt zu kommen. Gute Formatierung macht keine schlechte Arbeit gut (die inhaltlichen Fehler fallen eher umso deutlicher auf), aber schlechte Formatierung macht eine gute Arbeit schlechter.
- **Professionell und offen Kommunizieren:** Weder sich selbst kreuzigen noch Fehler verbergen, sondern berichten, was gemacht wurde und mit Fehlern konstruktiv umgehen.

Sprache: Stil

- **Kein Ich, keine Hero's Journey:** Es ist — im Prinzip — egal, wer die Arbeit durchgeführt hat (zumindest für die Qualität der Arbeit, nicht für die Bewertung Ihrer Leistung). Was überzeugen muss ist das Vorgehen, die Belege und Argumentation. Entsprechend stellen Sie das Vorgehen neutral dar ohne auf sich selbst zu verweisen (eher passiv verwenden). Ausnahmen sind u.a. in Danksagung, Widmung, Eidesstattliche Erklärung.
- **Meta vermeiden:** Sie müssen an vielen Stellen darauf hinweisen was kommt (z. B. zu Beginn eines jeden Kapitels). Reißen Sie den Leser aber dabei nicht aus dem Text. Mental ist der Leser dabei, ihre Arbeit zu beobachten, was Sie konkret getan haben. Wenn Sie ihm jetzt sagen "In diesem Kapitel ..." dann ziehen Sie ihn aus dem Text und bringen ihn dazu, über den Text nachzudenken statt über das, was gemacht wurde. Bleiben Sie bei dem, was Sie gemacht haben, z. B. "Nachfolgend werden ...".
- **Kapitel \neq Unterkapitel \neq Abschnitt:** 1 ist ein Kapitel, 1.1 ein Unterkapitel, 1.1.1 ein Abschnitt, 1.1.1.1 existiert nicht. Wenn Sie noch mehr Einteilungen brauchen, dann verwenden Sie Abschnitte mit Fettdruck zu Beginn (wie in diesem Abschnitt, dann aber ohne die Bulletpoints).
- **Kein Bulletpoint-Text:** Bulletpoints sind nur an wenigen Stellen hilfreich, z. B. bei Aufzählungen. Sätze nie mit Bulletpoints aufzählen. Entweder die Sätze auf Stichworte reduzieren oder eine Tabelle draus machen.
- **Auch digital gedrucktes ist tot und macht nichts mehr:** Sie berichten was

Sie gemacht haben — außerhalb des Berichtes. Entsprechend nie schreiben, dass z. B. in der Analyse der Sachverhalt analysiert wird. Der Bericht macht von sich aus nichts. Sie stellen dort die Ergebnisse der Analyse dar.

- **Umgangssprache vermeiden, Hochgestochene Sprache vermeiden:** Weder Umgangssprache (Sozialpädagogensprache; "tut", "Das Ganze ...", "etwas für ihre Gesundheit zu machen") noch Hochgestochen (à la Philosophendeutsch) schreiben. Wissenschaftliche Sprache ist nach Alley (1996):

- präzise: sagen was man meint (richtige Wort, richtiges Detaillevel)
- klar: vermeiden Sachen zu sagen/implizieren, die man nicht meint, d.h. Ambiguität und unnötige Komplexität (v.a. in der Wahl der Wörter) vermeiden
- ehrlich: direkt und offen kommunizieren
- prägnant: jedes Wort sollte zählen
- bekannt/vertraut: neue Fakten in bekannten Kontext verankern
- flüssig: von Satz zu Satz, Absatz zu Absatz, ohne dass der Leser stolpert

Alley, M. (1996). *The Craft of Scientific Writing* (Vol. 3). Springer

Sprache: Zeitformen

- **In einem Bericht berichten Sie:** Entsprechend Vergangenheitsform verwenden. Sie berichten über eine abgeschlossene Arbeit, selbst wenn diese noch läuft als Sie es geschrieben haben, und selbst bei Zielen der Arbeit. Ausnahmen sind selten, z. B. bei den Ergebnissen ("Daten zeigen ..." — sie machen es ja noch) und Ausblick bzw. Vorschläge für die Zukunft.

Sprache: Absätze

- **Eine Sinneinheit = 1 Absatz:** Absätze behandeln immer einen Punkt, eine Sinneinheit. Wenn eine halbe Seite lang kein Absatz verwendet wird, liegt meist ein Problem vor.
- **Absätze sind immer länger als ein Satz:** Keine Einsatz-Stückel-Absätze. Einzige Ausnahme: Sie wollen, dass der Leser die komplette Aufmerksamkeit auf

diesen zentralen Satz lenkt. Das kann man 1-2 Mal in einer Arbeit machen.

Interne Verweise

- **Verweise statt Wiederholungen:** Üblicherweise braucht man einen Sachverhalt nur ein Mal zu beschreiben — dann verweist man an anderer Stelle auf den konkreten Abschnitt. Das ist auch der Grund für die Nummerierung — es erlaubt Ihnen, den Leser präzise zu den Punkt in der Arbeit zu schicken, an dem Sie auf den Sachverhalt eingehen. Nutzen Sie also "wie in der Kontextanalyse (2.4) beschrieben ...", dann weiß der Leser, wie lange er blättern muss.

Zitationen

- **APA verwenden:** American Psychological Association (7. Ausgabe) Stil verwenden. Gibt genug Informationsseiten dazu im Netz und Literaturmanager können diesen Stil üblicherweise.
- **Richtige Zitationen:** Falsche "ich füg die richtigen Zitationen später ein" Abgaben verbrennen Ihnen den Betreuer. Arbeiten Sie von Anfang an mit einem Literaturverwaltungsprogramm, in dem die verwendeten Quellen richtig eingetragen sind (nächster Punkt).
- **Autorennamen müssen genannt werden, aber nicht hervorheben:** Die Wissenschaft sollte keinen Personenkult kennen — das können gerne Religionen oder Ideologien übernehmen. Wer was herausgefunden hat, ist egal. Die Qualität der Arbeit zählt. Entsprechend nicht "Die Autoren xyz haben herausgefunden das ABC vorliegt" sondern "Da ABC vorliegt (Autoren Jahr) ...".
- **Kein Paper-Denglish:** Ja, im englischen heißt es Paper. Im Text sind es aber Artikel oder Konferenzbeiträge oder Buchkapitel oder was auch immer. Üblicherweise muss man den Typ auch nicht erwähnen (im Normalfall wurde man eh Artikel/Konferenzbeiträge zitieren). Bitte nicht so was wie "Im Paper von xyz ...", das klingt nach Möchtegern-coole Manta mit Fuchsschwanz Sprechweise. Einfach berichten wie die Befundlage ist, über Autorennamen und Jahr (meist in Klammern) belegen wo es herkommt (sonst ein Plagiat) und die Belege und Argumente in der Arbeit sprechen lassen.
- **Wörtliche Zitate nur wenn es nicht anders geht:** In den meisten Fällen geben Sie Befunde oder Argumente mit Ihren eigenen wieder (mit Quellenangabe).

Wörtliche Zitate braucht man nur in sehr seltenen Fällen. Z. B. treffende Aussage von Evaluationsteilnehmern, oder eine Definition, die man 1:1 so sagen muss.

- **Fußnoten vermeiden:** Entweder es ist wichtig genug, genannt zu werden, oder es ist so unwichtig, dass es raus kann. Fußnoten reißen den Leser aus dem Text. Einzige Ausnahme: Bei der ersten Verwendung des generische Maskulinums.
- **Es gibt mehr Plagiate als nur Quellen nicht angeben:** Es gibt z. B. Übersetzungsplagiate, bei denen Sie einen Text(teil) einfach auf deutsch übersetzen ohne die Quelle anzugeben, oder dass man einfach die Argumentstruktur und Quellenangaben aus einem anderen Text übernimmt (ohne die Quelle anzugeben). Sie schmücken sich dann mit fremden Federn und Blender sind selten willkommen.
- **Literaturverwaltungsprogram nutzen:** Literaturverwaltungsprogramm (z. B. Zotero) hilft extrem bei der richtigen Zitierung, aber ACHTUNG: Wenn die Angabe in Zotero fehlt, ist auch die automatische Generierung des APA Stils falsch! GIGO gilt auch hier.

Abbildungen und Tabellen

- Expliziter Verweis vom Text auf die Abbildung/Tabelle immer im Absatz vor der Abbildung/Tabelle (Leser stolpern über Abbildung/Tabelle, suchen dann nach oben nach mehr Informationen).
- Auf **Lesbarkeit** achten! Schriftgröße und Auflösung (bei Bildern) im Probedruck überprüfen!
- Für Schwarz/Weiß-Druck und für Farbfehlsichtige geeignet.
- Tabellen nach **APA Stil** (nur horizontale Linien und nur nach Header oder vor Footer).

Druck

- Arbeit einseitig drucken!
- **PDF Druck und Suchfunktion:** Zuerst als PDF drucken, dann nach "Fehler! Verweisquelle konnte nicht gefunden werden." suchen. Word bricht schon mal gerne die Verlinkungen und das sieht man erst im Druck! Generell PDF Dokument kritisch durchgehen und PDF auch zum Drucken im Copyshop verwenden!

- Nach dem Druck und vor dem Binden alle Seiten selbst sowie von einer anderen Person durchgehen (lassen).

Acknowledgement

Kurzer Dank an Personen, die Sie bei der Arbeit unterstützt haben. Z. B. inoffizielle Betreuer, Teilnehmer an den Evaluationen (nie namentlich nennen), Medientechnik, Sekretärin, etc. pp. — nur wenn Sie den Personen wirklich dankbar sind. (Ist nett aber für die Bewertung irrelevant.) Falls nicht verwendet diese Seite einfach entfernen.

Text ...

Kurzfassung

Abstract schon für Zwischenabgabe schreiben. Später kommen dann noch Sätze dazu, aber Grundgerüst steht.

Kein "Teaser" sondern eine kurze Zusammenfassung (das, was man braucht, um sich schnell einen Überblick zu verschaffen, ob es sich lohnt, die Arbeit zu lesen).

Inhalt umfasst die zentralen Punkte aller Kapitel, von Ziel/Fragestellung bis Ausblick.

Nie länger als diese eine Seite (inkl. Schlüsselwörter).

Text ...

Schlüsselwörter

Verwendete Literatur gibt Hinweise auf passende Stichwörter. Das sind die Suchbegriffe, die man bei einer Literatursuche verwenden würde.

Text ...

Abstract

Englische Version der Kurzfassung. Nicht einfach Google Translate oder DeepL verwenden. Trifft die Nuancen nicht und klingt z. T. nach Yoda.

Text ...

Keywords

Text ...

Contents

1	Introduction	1
1.1	Goals	3
1.2	State of the Art	4
1.3	Approach	8
1.4	Definition of Terms	8
2	Analysis	11
2.1	Data Sources	12
2.2	Problem Analysis	14
2.3	Task Analysis	15
2.4	User Analysis	15
2.5	Context Analysis	15
2.6	Organisational Analysis	16
2.7	Conclusion on the Analysis	16
3	Conception	18
3.1	Conceptual Approach	20
3.2	Use Cases	20
3.3	Functionalities	21
3.4	System Architecture	21
3.5	Interface Design	22
3.6	Conclusion on the Conception	23
4	Implementation	25
4.1	System Architecture Implementation	26
4.2	Interface Implementation	26
4.3	Conclusion on the Implementation	26
5	Dialogue Samples	28

6	Summative Evaluation	29
6.1	Goal	29
6.2	Methods	29
6.2.1	Design	30
6.2.2	Participants	30
6.2.3	Setting and Instruments	30
6.2.4	Procedure	31
6.3	Results	32
6.4	Discussion	33
6.5	Conclusion on the Evaluation	33
7	Summary and Outlook	34
7.1	Summary	34
7.2	Outstanding Issues	34
7.3	Outlook	35
7.4	Final Conclusion	35
	List of Figures	36
	List of Tables	37
	Sources	38
	References	38
	Websites	40
	Software	40
	Abbreviations	41
	Glossary	42
	Appendices	43
	Appendix A: Inhalt der DVD	43
	Appendix B: Anhangstitel	44
	Assertion under Oath	45

1 Introduction

The use of modern Artificial Intelligence (**AI**) techniques is pervasive and can be found in many fields of application, such as digital image processing, search engines and speech recognition (European Commission, 2020). Other application fields, such as medical diagnosis systems, cannot benefit as easily from AI-based technology compared to recreational domains. This impediment stems oftentimes from the AI being a "black box". In consequence humans struggle to understand such AI-systems and their output, leading to trust and compliance issues (Adadi & Berrada, 2018). These issues are further enhanced in the medical context where decisions, possibly based on AI, can have severe consequences for users, especially patients.

An example for medical human-AI-interaction is the image-based recognition of Deep Vein Thrombosis (**DVT**) with real time AI support for medical professionals by *ThinkSono*. The system leverages AI to guide the user through the current gold-standard diagnosis, a compression ultrasound examination, so that it enables any healthcare professional to detect DVT (ThinkSono, 2021). Closely related in this context is the interdisciplinary research project *CoCoAI*, which aims to explore psychological, ethical and technological implications of human-centered, AI-based applications in the DVT diagnosis and beyond (CoCoAI, 2021).

When AI-based systems are used in high risk application contexts, such as medical diagnosis, the aspects of explainability, interpretability and trustworthiness become a primary concern for adoption and use of said system. Ribeiro et al. (2016) already explored the importance of explainability and trust in AI-based systems and postulated that AI systems will not be used if the users have no trust in the model or the results. Even though many machine learning algorithms score high on standard performance metrics, such as precision, recall or Area Under the Receiver Operating Characteristics (**AUROC**), user-facing performance may be way worse (Gordon et al., 2021). Understanding the AI's underlying machine learning (**ML**) model and its predictions is an important step for assessing trust and facilitating effective interaction (Ribeiro et al.,

2016). Recent technological advances are realized by *Clearbox AI*, with the focus on trustworthy AI by implementing an AI model assessment (Clearbox AI, 2021b; European Commission, 2021). The model assessment can help model owners to identify robustness issues, potential undesired behaviour, and explain errors and uncertainties regarding the model predictions (Clearbox AI, 2021a).

Trust in AI systems is primarily induced by the users' understanding and the general interpretability of the machine learning model and their predictions (Ras et al., 2018; Ribeiro et al., 2016). The wide array of different possible user groups and the complex constructs of understandability and trust demands for a human-centric approach in designing AI assessment systems. Because of the inherent complexity of non-linear machine learning models, especially Deep Neural Networks (**DNNs**) for image processing, suitable visualization and communication techniques are non-trivial. Additionally to the complex models for image classification, the input data is also more complex as it is unstructured. Non-linear neural networks and unstructured data provide additional challenges for Explainable Artificial Intelligence (**XAI**), as described in Keane and Kenny (2019). XAI is a research field that studies how AI decisions and data driving those decisions can be explained to people in order to provide transparency, enable assessment of accountability, demonstrate fairness, or facilitate understanding (Arrieta et al., 2019). XAI plays an important role in the acceptance and finally in the usage of AI-based technology. This is further underlined in the medical context where public authorities set strict regulations on the usage of technological systems and ethic concerns have to be thoughtfully addressed.

In the context of a image-based medical diagnosis system, it is important that the responsible stakeholders, such as medical practitioners, specialized doctors and clinic managements, are enabled to make informed decisions on the usage of AI-based technology, even though their expertise in machine learning and data science is expected to be low. The stakeholders' trust in this system is a primary factor for the widespread adoption of said technology for real life applications. Therefore, increasing the understanding of the AI model and finding an optimal trust level in the predictions by designing human-centric explanation techniques within the AI model assessment system is a main goal of this work. Additionally it is conceivable that authorities will instantiate auditors for AI-based systems in medical contexts. Having a comprehensible and scientifically proven assessment system could be a big step in the approval and adoption of said system.

1.1 Goals

The users understanding of the AI model and trust in the model are highly essential as pointed out by Knapič et al. (2021). This holds especially true for medical applications where re-traceable results have to be provided and people acting on these results bear great responsibility. To facilitate understanding and trust the machine learning model has to be interpretable and explainable. In the context of Convolutional Neural Networks (CNNs) interpretability of models can pose a significant concern because of their inherent complexity. Explanations of AI-models can provide insights on the machine's decision process and therefore generate user understanding. This can lead to the model being more interpretable by humans.

Assessing the suitability and performance of a CNN for a specific task by applying standard performance evaluation metrics is problematic, since these can be oblivious to distinguishing the diverse problem solving behaviors of a neural network (Lapuschkin et al., 2019). Lapuschkin et al. (2016), Ribeiro et al. (2018), Samek et al. (2021) give an overview on the technical foundations of XAI and a presentation of practical methods, which will be used in conjunction with human-centric design to explore and evaluate suitable and efficient methods to explain a model's classification.

The goal of this thesis is to design and develop and evaluate interactive AI-assessment-system artifacts for medical professionals and machine learning specialists in a human-centric fashion to facilitate understandability and trustworthiness of AI-models. Developing such a system, with human concerns in focus, leads to following research questions:

- Q1: How is the stakeholders' (medical professionals, clinic managements or data scientists) subjective information processing awareness linked to trust for a specific model and its predictions in the medical domain?
- Q2: How can different explanation techniques, ranging from perturbation based approaches to more model intrusive alternatives, increase trust in image classifier models and predictions?
- Q3: What are the most efficient methods to explain and possibly optimize trust levels in image classification models?
- Q4: To what extend can structured metadata increase the stakeholder's understanding of a model's operational range and performance?

1.2 State of the Art

An AI-assessment-system is currently offered by Clearbox AI. The *AI Control Room* cloud platform enables users to assess, improve and validate ML models and data in accordance with the principles of Trustworthy AI (Clearbox AI, 2021b; European Commission, 2021). Clearbox AI (2021b) describes its AI-assessment-system as a "Deep Pre-production Analysis" tool:

"AI Control Room automatically generates a model assessment to help model owners to identify robustness issues, potential undesired behaviour, and explain errors and uncertainties regarding the model predictions."

Concretely the product enables users to perform following tasks for AI models working on tabular data:

Model behaviour validation: Validation metrics and plausible causes of error are clearly presented, potential limitations and irreducible uncertainty are identified and local explanations of the model behaviour are generated selecting representative points in the dataset.

Synthetic data generation: A generative model can be used to create synthetic data points that preserve the statistical properties of the original dataset. These points can augment the original training set to improve generalization, to increase model robustness, and to oversample specific labels when in the presence of unbalanced data.

Data-centric analysis: Generative models perform a probabilistic analysis of the underlying data allowing for robust outliers detection and uncertainty analysis. This information can help you to evaluate data quality.

Centralised tracking system: AI Control Room acts as a centralised tracking system to store lineage, versioning, and metadata of your datasets and models. Assessments generated are securely persisted along with models and datasets.

Besides general information and standard metrics (see Figure 1.1) the assessment system offers varied insights into different aspects of the machine learning model: Figure 1.2 shows graphs of training and validation precision, recall and calibration, while Figure 1.3 shows the models strong points and limitations by analyzing the feature distribution in

the data. Furthermore, the second half of the model assessment focuses more on the interpretability aspect of machine learning: Figure 1.4 shows a confusion matrix of possible classification results, which is then extended by example data points, chosen by the assessment system (see Figure 1.5). These examples can then be further explored to generate understanding of the models inner workings by applying an attribution based explanation technique combined with a decision rule explanation.

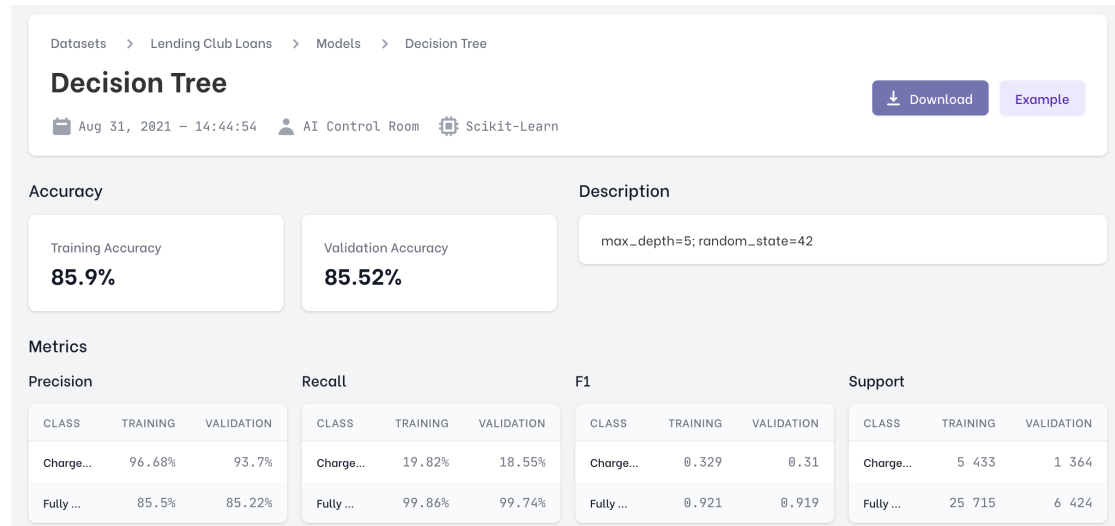


Figure 1.1: AI Control Room - Model Assessment Overview with Standard Metrics

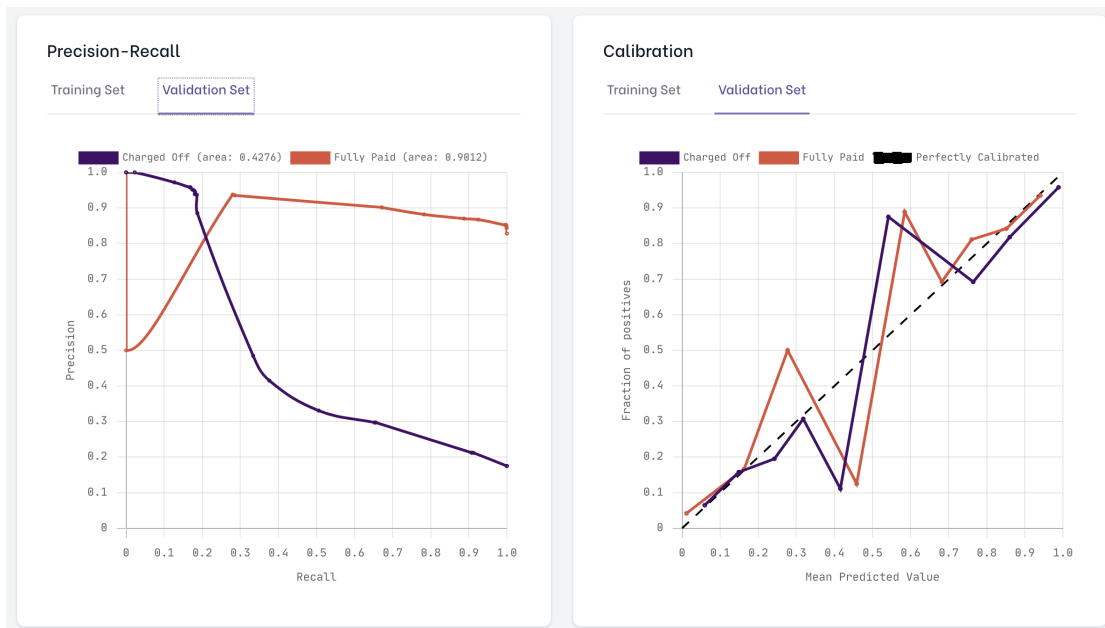


Figure 1.2: AI Control Room - Precision-Recall and Calibration Graphs

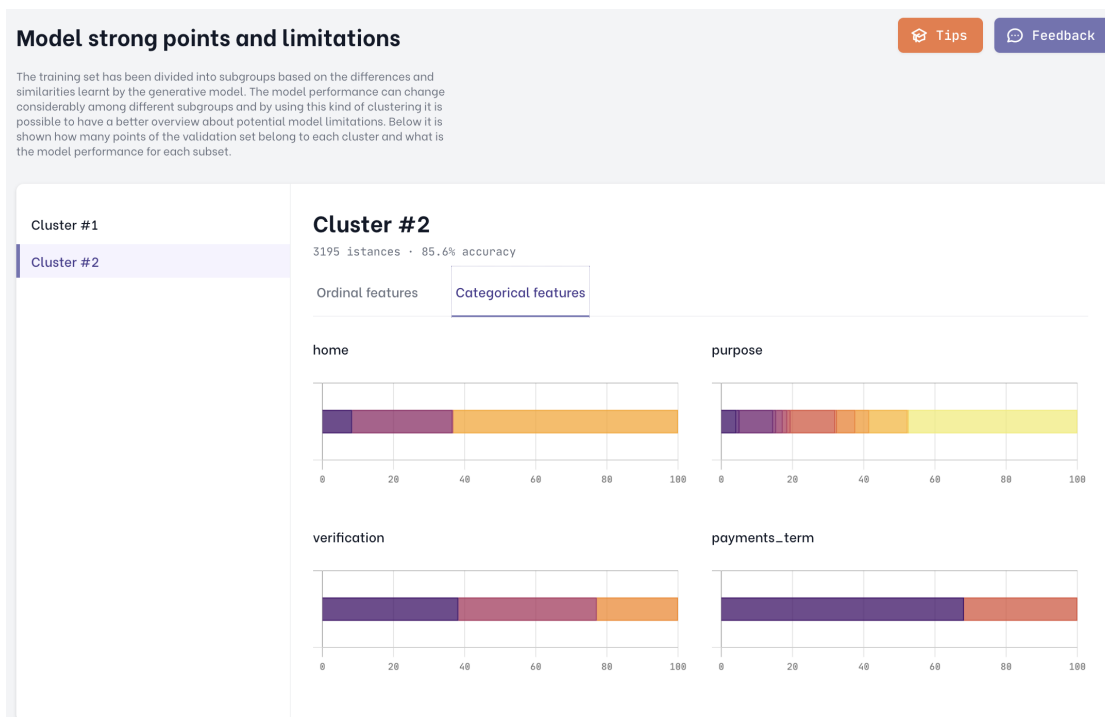


Figure 1.3: AI Control Room - Model String Points and Limitations

Interpretability assessment

[Tips](#)[Feedback](#)

By clicking on an element of the confusion matrix, one can see a selection of the most representative points for that element. These points have been chosen by our tool by selecting the most diverse and representative instances from each of the clusters defined above.

Each cell of the dataframe has been colored using the value of the attribution model generated for the respective point. By clicking on each point one can see its local explanation in the form of decision rule and corresponding local attribution value.

		Predicted	
		Charged Off	Fully Paid
Actual	Charged Off	253 3.25%	1111 14.27%
	Fully Paid	17 0.22%	6407 82.27%

Representative examples predicted as **Charged Off**, while ground truth was **Charged Off**

Figure 1.4: AI Control Room - Interpretability Assessment

HOME	GRADE	PURPOSE	INQUIRIES	DELINQUENCY	LOAN_AMOUNT	BANKRUPTCIES	FICO_AVERAGE	VERIFICATION
Mortgage	2	debt_consolidation	4	0	10000	0	732	Not Verified
Mortgage	4	other	5	5	8000	0	667	Source Verified
Rent	6	other	5	0	18000	0	677	Verified
Mortgage	5	credit_card	8	0	12500	0	677	Verified
Rent	7	debt_consolidation	3	0	5875	0	642	Not Verified
Mortgage	3	credit_card	5	0	19000	0	722	Verified

Figure 1.5: AI Control Room - Example Data

Explanation

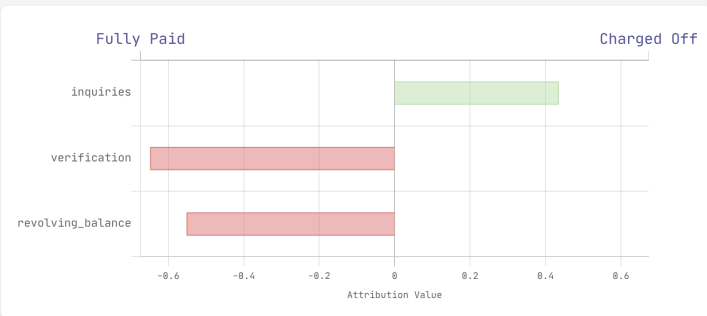
Prediction

CLASS	PROBABILITY
Fully Paid	0.0053
Charged Off	0.9947

Explanation Confidence ?

99.92%

Attribution chart ?



Decision rule ?

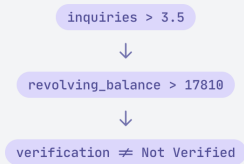


Figure 1.6: AI Control Room - Prediction Explanation for Examples Data

1.3 Approach

As already described in the introduction of chapter 1, many machine learning algorithms score high on standard performance metrics, but user-facing performance may be way worse (Gordon et al., 2021). This issue is caused by real world applications being very dependent on the actual human-AI-interaction. Following this reasoning, it lends itself to utilize a human-centered design process for creating AI-assessment systems. Figure 1.7 shows a standardized process of human-centered design, which was applied in this thesis to conceptualize, implement and evaluate assessment system artifacts. The key take-away is the inclusion of human aspects in all stages of the process. Research on evaluation of AI explanations revealed that there is a big gap between the perceived and actual usefulness of explanations, as described by Ras et al. (2021). This further underlines the need for a human-centered approach in designing AI-assessment-systems.

The thesis' structure will reflect the human-centered approach, which is visualized in Figure 1.7: As already alluded in chapter 1, chapter 2 is about understanding and setting the context of use. The following chapter 3 deals with determining the requirements for use by conducting literature research and user interviews. Based on the established requirements the development of solutions will be described in chapter 4, while chapter 6 is about the evaluation of the solutions. This general process is embedded in a iterative loop, where intermediate results are evaluated against the requirements and subject to change.

1.4 Definition of Terms

Text ...

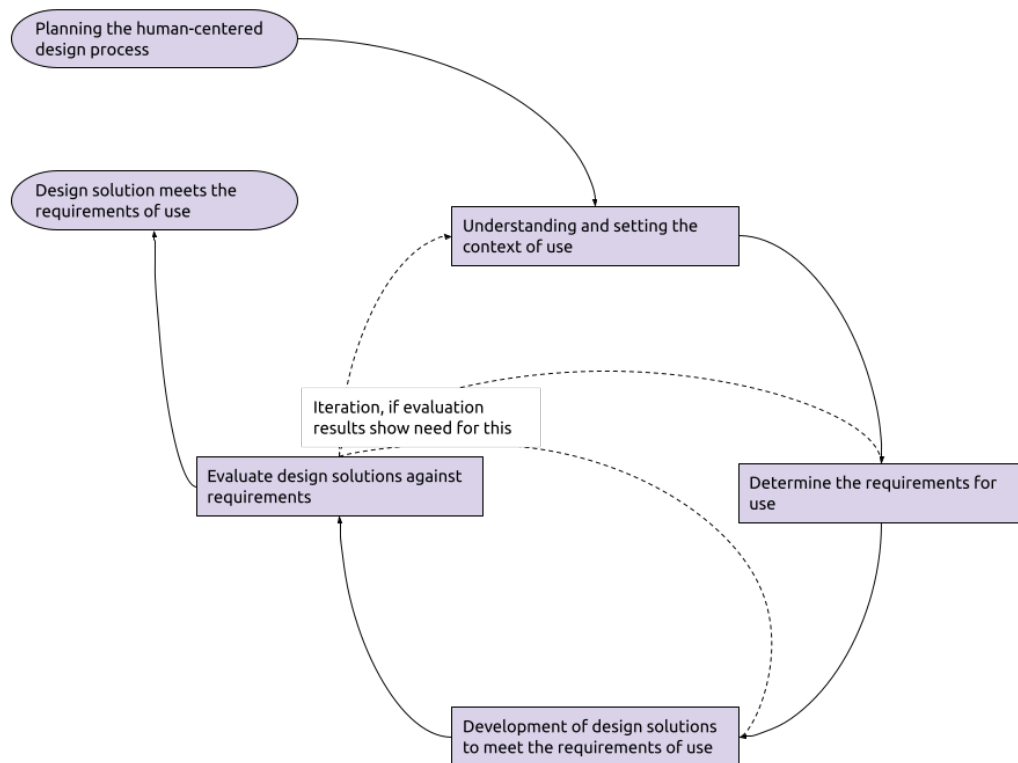


Figure 1.7: Human-centered Design Process (*DIN EN ISO 9241-210*, 2011)

STOP

Halten Sie an dieser Stelle für einen Moment inne. Sie wissen, was Sie erreichen wollen, haben aber die Analyse noch nicht begonnen oder noch nicht abgeschlossen. Halten Sie auf einem Blatt Papier fest, wie Sie sich die Lösung für das Problem vorstellen. Wenn

Sie mehrere Lösungen haben, um so besser. Legen Sie das Papier dann beiseite.

Schauen Sie es auch erst einmal nicht mehr an.

Sie brauchen es erst **nach der ersten Iteration der Konzeption** wieder.

2 Analysis

Über das Verstehen und Festlegen des Nutzungskontexts zum Festlegen der Nutzungsanforderungen: Das Ziel der Analyse besteht aus den ersten beiden Iterationsschritten des menschenzentrierten Gestaltungsprozesses: das **Verstehen und Festlegen des Nutzungskontexts**, was am Ende der Analyse im **Festlegen der Nutzungsanforderungen** mündet.

Anforderungen verstehen, nicht schon Lösungen entwickeln: Wenn Sie während der Analyse gute Ideen haben, dann diese sofort notieren. Aber in die Analyse kommt nur das, was für das Verstehen und Festlegen der Nutzungskontextes bzw. der Festlegung der Nutzungsanforderungen relevant ist. Ihre Ideen für Lösungen gehören in die Konzeption. Schließlich ist Ihre Lösung vermutlich nur eine von vielen und jemand, der nur bis zum Ende der Analyse liest, um dann selbst Lösungen zu konzipieren, sollte nicht von Ihren Lösungsideen beeinflusst werden.

Advance Organizer: An dieser Stelle (nach 2, vor 2.1) erfolgt eine kurze Übersicht, was den Leser in den Unterabschnitten erwartet. Das gibt dem Leser einen Überblick, wie Sie in Ihrer Analyse vorgegangen sind und erlaubt ihm zu interessanten Abschnitten zu springen. Gilt auch für alle weiteren Kapitel.

Struktur der Analyse: Die Reihenfolge der Analyse-Unterkapitel (Problemanalyse, Aufgabenanalyse, Benutzeranalyse, Kontextanalyse, Organisationsanalyse) muss eine logische Reihenfolge ergeben. Häufige Reihenfolgen (nach — falls sinnvoll — 2.1 Datenquellen) sind:

- **Problem-/Aufgabenanalyse => Benutzeranalyse => Kontextanalyse => Organisationsanalyse:** "das ist das Problem/Aufgaben, das die Benutzer haben, die es unter den räumlichen und zeitlichen Bedingungen unter den organisationalen Rahmenbedingungen lösen müssen".

- **Benutzeranalyse => Problem-/Aufgabenanalyse => Kontextanalyse => Organisationsanalyse:** "die Benutzer sind gekennzeichnet durch (z. B. Patienten mit Depression), die aufgrund ihrer Eigenschaften die folgenden Probleme/Aufgaben haben, die sie unter den räumlichen und zeitlichen Bedingungen unter den organisationalen Rahmenbedingungen lösen müssen".
- **Organisationsanalyse => Problem-/Aufgabenanalyse => Benutzeranalyse => Kontextanalyse:** Organisationale Rahmenbedingungen haben sich verändert (z. B. zunehmende Digitalisierung), das führt zu neuen Problemen/Aufgaben bei den Benutzern, die es unter den räumlichen und zeitlichen Kontextbedingungen lösen müssen.

Wenn Sie die richtige argumentative Reihenfolge haben, können Sie die Analyse (sehr) abstrakt in einem Satz zusammenfassen (z. B. **Benutzeranalyse => Aufgabenanalyse => Problemanalyse => Kontextanalyse => Organisationsanalyse:** "Depressionspatienten sind durch verzerrte Erinnerungen gekennzeichnet, um die Therapie zu unterstützen, müssen sie Ihre Emotionen protokollieren wenn diese auftreten, das ist allerdings ziemlich aufwändig und eine Lösung dafür muss nicht nur gebrauchstauglich sein, sondern auch unterwegs funktionieren und sich der Kommunikation mit dem Therapeuten und seiner Infrastruktur anpassen."). In den Unterkapiteln tragen Sie dann konkret zusammen, was für die menschenzentrierte Gestaltung bezüglich Benutzer, Aufgabe, Problem, Kontext, und Organisation zu berücksichtigen ist. Sie tragen damit die Legobausteine für ihre Konzeption zusammen.

Häufig wird die Analyse allerdings dadurch erschwert, dass Sie unterschiedliche Probleme/Aufgaben haben, die von unterschiedlichen Benutzer(klassen) unter unterschiedlichen räumlichen und zeitlichen Bedingungen unter unterschiedlichen organisationalen Bedingungen ausgeführt werden (vgl. im Gegensatz zu der sehr speziellen eindeutige App für Depressionspatienten im klinischen Kontext z. B. ein allgemeines Programm wie Microsoft Word, das von einer sehr heterogenen Nutzergruppen verwendet wird).

2.1 Data Sources

In Ihre Analyse verwenden Sie wahrscheinlich die selben Datenquellen (z. B. wissenschaftliche Literatur, bestehende Apps, Experteninterviews, Nutzerbefragungen, etc.)

in mehreren Unterkapiteln. Hier können Sie diese Quellen kurz beschreiben um sie in den folgenden Unterkapiteln verwenden zu können. Das macht nur dann Sinn, wenn Sie die Daten auch wirklich in mehreren Unterkapiteln verwenden. Wenn Sie z. B. bestehende Apps nur in der Problemanalyse verwendet haben, dann das Vorgehen auch nur dort beschreiben (wäre allerdings was merkwürdig, wenn Ihnen die Apps nicht auch was über die anvisierten Nutzer sagen würde).

Generell müssen Sie die Quellen so beschreiben, dass jemand anders — wenn er das Vorgehen wiederholen würde — zum selben Ergebnis kommen würde (Transparenz / Nachvollziehbarkeit!).

Quellen und Informationen können z. B. sein:

- **Nutzerinterviews:** Beschreiben wie Sie die Personengruppen bestimmt haben, wie sie die Personen ausgewählt haben, welche Fragen Sie gestellt haben (nur Oberkategorien, kompletten Interviewleitfaden im Anhang). Bei den oft wenigen Personen (ca. 10) eine Tabelle mit ID, Alter, Geschlecht, Merkmal warum die Person für das Interview relevant war (z. B. Berufsbezeichnung, Berufserfahrung, etc.). Evtl. statistische Kennwerte (Mittelwert) aus relevanten Fragebögen (z. B. ATI). Im späteren Text verwenden Sie dann die ID, was dem Leser erlaubt zu sehen, ob mehrere Aussagen auf die selbe Person zurückgehen.
- **Umfragen:** Wann wurde die Umfrage worüber wie verbreitet? Welche Fragen wurden gestellt (nur Oberkategorien hier nennen, für kompletten Fragebogen auf den Anhang verweisen). Wie viele Personen haben teilgenommen? Welche zentralen Eigenschaften haben die Personen? Üblicherweise Mittelwert und Standardabweichung von Alter, relevanten Skalenwerten wie z. B. ATI. Bei Geschlecht Häufigkeit von männlich, weiblich, und divers. Für weitere relevante Eignungskriterien (z. B. Berufsbezeichnung, Berufserfahrung) Häufigkeiten bzw. Mittelwert und Standardabweichung angeben.
- **Wissenschaftliche Literatur:** Wo (Suchbegriffe) haben Sie wann mit welchen Suchbegriffen gesucht? Oft relevante Datenbanken sind z. B. Scopus, ScienceDirect, Google Scholar, ACM Digital Library, oder die GI Digital Library. Je nach Thema zusätzlich fachspezifische Datenbanken verwenden (z. B. für Psychologie, Medizin, etc.).
- **Bestehende Anwendungen:** Oft bietet sich eine Google Suche oder eine Suche

in den verschiedenen App-Stores (iOS, Android, Mac App Store, etc.) an. Auch hier angeben, wo Sie wann und mit welchen Keywords gesucht haben.

Genaue Angaben bei der Recherche (wissenschaftliche Literatur, Bestehende Anwendungen) schützen Sie vor der Frage: "Warum haben Sie den Artikel X oder Anwendung Y nicht berücksichtigt?" — sofern Ihre Recherche gut von Ihnen durchgeführt und dokumentiert wurde.

Insbesondere die Literaturrecherche wird häufig unterschätzt. Es macht einfach mehr Spaß, eine Lösung, die man im Kopf hat, zu skizzieren, anstatt zunächst einmal sicher zu gehen, dass man den Nutzungskontext überhaupt verstanden hat. Meist erfindet man das Rad damit allerdings nur neu oder kommt zu Lösungen, die nett aus-sehen, aber für den konkreten Kontext unbrauchbar sind.

Achtung: In diesem Abschnitt geht es darum, wo Ihre Daten für die Folgeabschnitte in der Analyse herkommen. Es geht explizit hier nicht darum, was konkret rausgekommen ist. Zum Beispiel kommen Informationen über die Benutzer — jenseits von der Eignung für die Befragung — in die Benutzeranalyse! Dort verweisen Sie dann auf diesen Abschnitt bzw. auf den Unterabschnitt Nutzerinterviews oder Nutzerbefragung. Vorteil: Wenn Sie Ergebnisse aus diesen Befragungen auch in anderen Unterabschnitten der Analyse verwenden, können Sie dort ebenfalls auf den entsprechenden Abschnitt hier verweisen.

Text ...

2.2 Problem Analysis

Problem- und Aufgabenanalyse wird häufig kombiniert, können aber auch in zwei separate Unterkapitel aufgeteilt werden.

Problemanalyse: Welche(s) Problem(e) adressieren Sie? Machen Sie es konkret. Verwenden Sie empirische Daten, z. B. über die Häufigkeit des Problems. Machen Sie es plastisch mit einem typischen Szenario. Weisen Sie in der weiteren Arbeit explizit darauf hin, wie Ihre Entwicklung diese Probleme adressiert.

Text ...

2.3 Task Analysis

Aufgabenanalyse: Welche Aufgaben sollen die Benutzer/innen (besser) ausführen können? Wodurch sind diese gekennzeichnet? Bei kleineren klar definierten Aufgaben bieten sich z. B. HTAs an. Nehmen Sie in der weiteren Arbeit explizit Bezug auf die Aufgaben und die durch die Entwicklung neu vorhandenen Möglichkeiten bzw. bessere Durchführung.

Text ...

2.4 User Analysis

Wer sind die Benutzer/innen Ihrer Entwicklung? Beschreiben Sie die Benutzergruppen oder Benutzerklassen. In **seltenen(!)** Fällen können sich Personas eignen (sind eher ein Kommunikationstool im Team). Die Zielgruppe müssen Sie im Verlauf Ihrer weiteren Arbeit immer wieder explizit einbinden — nutzen Sie sie um Ihre Gestaltungslösungen immer wieder an deren Eigenschaften zu messen. In Konzeption, Realisierung, und Evaluation muss deutlich werden, dass Sie spezifisch für diese Personen und ihre Besonderheiten entwickelt haben.

Teilen Sie unterschiedliche Benutzerklassen in Unterabschnitte auf, z. B. "2.4.1 Verwaltungsmitarbeiter", "2.4.2 Bürger". Es sollten nicht mehr als ca. drei Benutzerklassen bedient werden.

Text ...

2.5 Context Analysis

Der Kontext bezieht sich auf den **räumlichen** und **zeitlichen Kontext**, in dem Ihre Entwicklung eingesetzt wird (z. B. **Informations-Website:** auf Desktop-Computern, Notebooks und Mobilgeräten sowohl zu Hause als auch unterwegs und zu jeder Uhrzeit vs. **Sachbearbeitungssoftware in der öffentlichen Verwaltung:** auf Arbeitsplatz-Computern, im Büro, zwischen 9 und 17 Uhr) Aus diesen Kontexten ergeben sich unter-

schiedliche Nutzungsanforderungen, die für die Konzeption relevant sind. Beziehen Sie diese Informationen bei der weiteren Entwicklung explizit mit ein.

Wenn Sie Aussagen über den Kontext machen, dann legen Sie auch Ihre Quellen offen — woher wissen Sie das? Z. B. aus Interviews, der Literatur, etc.

Text ...

2.6 Organisational Analysis

Hier geht es um den **organisationalen Kontext**, also z. B. um Hierarchien, Abhängigkeiten zwischen den Nutzern/innen und anderen Personen/Organisationen, Rechte in Unternehmen, und andere Rahmenbedingungen, die Sie berücksichtigen müssen.

Nehmen Sie als Beispiel Ihre Situation als Studierende bei der Entwicklung einer Studiengangs-App — Sie sind im Kontext Universität mit Dozierenden, Universitätsleitung, Studiengangs-Koordination, entsprechenden Studien- und Prüfungsordnungen etc. unterwegs. Das müssen Sie bei der Entwicklung berücksichtigen.

Oder wenn es um eine Software für eine Klinik geht — und die IT den Einsatz von Java verbietet. Wenn Sie das nicht in Erfahrung gebracht haben und dann eine Java Lösung entwickeln, bekommen Sie erst viel zu spät heraus, dass Ihre Entwicklung nicht verwendet werden kann (yup, realer Fall — für die Klinik-IT ist das Java-Verbot normal und nicht der Rede wert, die Nutzer wissen davon oft nicht — oder gar was Java ist). BTW, versuchen Sie nicht, eine Organisation zu ändern — Organisationsentwicklung ist grausam und dafür werden sie nicht bezahlt.

Auch hier: Nehmen Sie auf diese Punkte explizit in der weiteren Entwicklung Bezug!

Text ...

2.7 Conclusion on the Analysis

Stellen Sie zum Ende der Analyse in einem eigenen Fazit mit eigener Gewichtung die zentralen Inhalte der Analyse kurz dar. Zur Unterstützung kann eine Tabelle mit einer Priorisierung der Nutzungsanforderungen verwendet werden. Sie legen **auf Basis der Analyse** begründet fest, welche Nutzungsanforderungen:

1. absolut notwendig erfüllt werden müssen (müssen Sie lösen, sonst kann das Ziel der Anwendung nicht erreicht werden, d.h. sie ist nicht effektiv),
2. wichtig zu erfüllen sind (hoher Einfluss auf Effizienz, Zufriedenstellung, Erlernbarkeit),
3. was "nice-to-have" ist (geringer Einfluss auf Effizienz, Zufriedenstellung, Erlernbarkeit), und
4. was eher unwichtig ist, aber die Anwendung abrundet (Kür).

Die Anforderungen mit **1 müssen** Sie erfüllen, die Anforderungen mit **2 sollten** Sie erfüllen, die Anforderungen mit **3 können** Sie erfüllen, und die mit **4 können** Sie erfüllen, wenn Sie wider erwarten am Ende der Entwicklungsphase noch was Zeit haben.

Falls bei der Festlegung der Nutzungsanforderungen viel Entscheidungsspielraum vorliegt — z. B. weil die Analyse keine klaren Prioritäten nahelegt, dann können Sie die Priorisierung auch zu Beginn der Konzeption darstellen. In der Konzeption haben Sie mehr gestalterische Freiheiten als in der Analyse. Ein eigenes Fazit sollten Sie am Ende der Analyse aber in jedem Fall schreiben.

Text ...

3 Conception

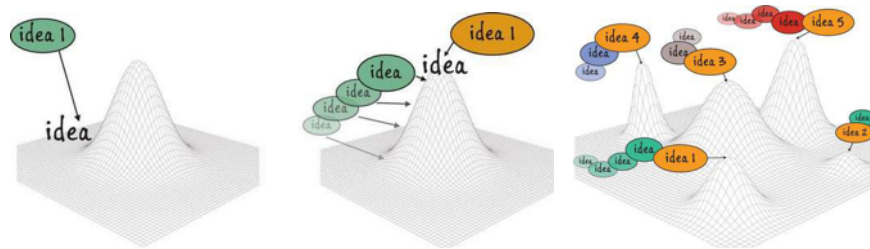
Die **Konzeption muss klar aus der Analyse folgen**. Sie können und sollten kreative Lösungen erarbeiten, aber Sie müssen dabei wirklich kreativ sein. Sprich: **Absichtlich etwas Erschaffen, was neu und nützlich ist**. Machen Sie deutlich wie Ihre kreative Lösung die Punkte aus der Analyse aufgreift und löst. Also nicht: "Jetzt habe ich die Pflicht-Analyse gemacht, kann's abhaken, und mich frei entfalten." Die Analyse liefert die **Legobausteine** und den **Rahmen** (Bodenplatte) für die Konzeption — die müssen Sie nutzen. Sie können natürlich iterativ die ihr Verständnis des Nutzungskontextes erweitern und die Nutzungsanforderungen updaten!

Stellen Sie zu Beginn der Konzeption einen **Bezug zur Analyse** her. Geben Sie dann einen Überblick über das Konzeptionskapitel. Wie sind Sie bei der Konzeption vorgegangen? Welche Aspekte (z. B. Funktionalitäten, Systemarchitektur, Interface-Design, etc.) berücksichtigen Sie? Verweise auf die Abschnitte (Abschnittsnummern verwenden!).

Zentral für die Konzeption ist die **formative Evaluation der Gestaltungslösung anhand der Anforderungen**. Sie wissen ja aus der Analyse wie die Nutzungsanforderungen sind — entsprechend immer zuerst selbst die Lösungen damit bewerten (= evaluieren). Dabei helfen auch Benutzerklassen oder (wenn sinnvoll) Personas. Aber frühzeitig die Zielgruppe hinzuziehen, damit diese Ihnen Feedback zu Ihren bisherigen Konzeptionen gibt. Sie brauchen oft mehrere Iterationen bis Sie eine gebrauchstaugliche Lösung herausarbeiten können.

Zuerst "Getting the Right Design", dann erst "Get the Design Right": Viele Arbeiten sind zu unkreativ. Das heißt nicht anders um anders sein zu wollen (wir sind nicht im Marketing), sondern etwas entwickeln, was neu und nützlich ist. Häufig wird mit der ersten Idee gestartet und diese iterativ verbessert. Das Ergebnis ist, dass man sich einen Hügel hocharbeitet (Abbildung linke Seite). Das führt zu einem "hill climbing problem" — man kommt evtl. oben an, aber übersieht, dass es andere Lösungen gibt,

die einen vielleicht weiter führen (Abbildung rechte Seite).



"Getting the Design Right" (links) vs. "Getting the Right Design" (rechts). Abbildungen aus Greenberg et al. (2012).

Stattdessen erst unterschiedliche Ideen mit sehr geringen Änderungskosten (= billige und grottige Papierskizzen) explorieren, dann erst für die beste Idee entscheiden (auf Basis der Nutzungsanforderungen, Ihren Fertigkeiten und zur Verfügung stehenden Zeit!)

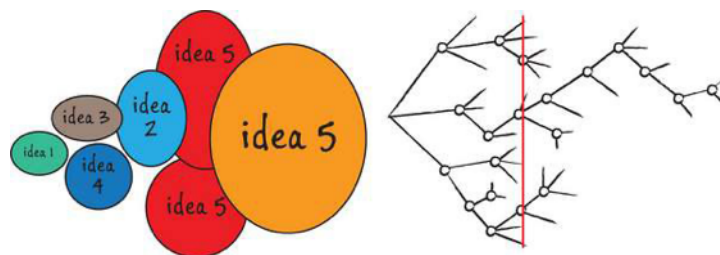


Abbildung aus Greenberg et al. (2012).

Die Konzeption **kann unterschiedlich aufgebaut sein** und **nicht alle Unterkapitel sind bei jeder Entwicklung relevant**. Die Struktur muss Sinn ergeben — im Zweifelsfall mit dem Betreuer abstimmen. Die Unterkapitel Konzeptionsvorgehen, Anwendungsfälle (Use-Cases) und Funktionalitäten wird man in den meisten Fällen benötigen. **Mögliche Reihenfolgen danach sind:**

- **Iterationsweise:** Bietet sich an, wenn Sie nur einen Aspekt (z. B. nur das Interface) entwickeln. Sie beschreiben in eigenen Unterkapiteln die jeweiligen Iterationen samt formativer Evaluation, aus der sich dann die nächste Iteration ableitet.
- **Strukturell gegliedert:** Aufteilung in Unterkapitel (Systemarchitektur, Interface, etc.) und Beschreibung der Iterationen dieser in den jeweiligen Unterkapiteln.

Text ...

3.1 Conceptual Approach

Die Konzeption verläuft — noch stärker als die Analyse — iterativ. Die Unterkapitel sind allerdings linear strukturiert. Vergleichbar mit den Datenquellen der Analyse bietet sich das erste Unterkapitel an, das Vorgehen und die eingesetzten Methoden zu beschreiben (deren Ergebnisse sich meist in mehreren Unterkapitel niederschlagen).

Methoden können z. B. sein:

- **Analyseworkshop:** Nützlich um die Ergebnisse der Analyse zu strukturieren und für die Konzeption aufzubereiten (vgl. Kumar (2013, S. 190-192))
- **Design-Sprints:** Stärker für die Konzeption (vgl. Knapp et al. (2016))
- **Workshops mit der Zielgruppe:** Generell um Ideen partizipativ zu entwickeln und formativ zu evaluieren.

Auch hier: Vorgehen beschreiben, Teilnehmer mit Alter (M, SD), Geschlecht (m/w/d) sowie relevanten Eigenschaften (warum stehen sie für die Zielgruppe?).

In den folgenden Unterkapitel können Sie dann auf 3.1 verweisen.

Text ...

3.2 Use Cases

Use Cases bieten sich an, die Ergebnisse aus der Analyse zusammenzufassen und in einen oder mehrere konkrete Anwendungsfälle zu übertragen. Die Tabelle mit den Nutzungsanforderungen sollte hierbei helfen. Wie sieht die Situation aus, wenn Ihr System eingesetzt wird? Bei den Anwendungsfällen noch nicht zu spezifisch werden. Es geht primär darum zu beschreiben, was die Person macht, z. B. veränderte Arbeitsabläufe, neue Funktionen, etc., noch nicht wie genau dies (technisch, womit realisiert) umgesetzt wird.

Text ...

3.3 Functionalities

Geben Sie einen Überblick über die geplanten Funktionalitäten des Systems. Dazu können Sie die Anwendungsfälle der Konzeption sowie die Tabelle mit den Nutzungsanforderungen aus der Analyse aufgreifen und den Anforderungen mögliche Funktionalitäten zuordnen. Priorisieren Sie dann diese Funktionalitäten — bezüglich Auswirkungen auf die Gebrauchstauglichkeit des Systems und Umsetzbarkeit dieser Funktionen.

Text ...

3.4 System Architecture

Mit den Funktionalitäten gehen bestimmte Anforderungen an das zugrundeliegende System einher. Stellen Sie hier dar, welche Lösungen es für die Systemarchitektur gibt, welche Aufteilungen sich eignen (z. B. MVC) und wie man es umsetzen könnte. Je nach Detailgrad bieten sich u.a. UML-Diagramme an.

Begründete Auswahl: Nehmen Sie die Funktionalitäten als Grundlage, das gibt Ihnen vor, mit welchen möglichen Architekturen Sie diese Funktionalitäten umsetzen können. Falls es mehrere Möglichkeiten gibt, erläutern Sie diese und treffen eine begründete Auswahl auf Basis dessen Stärken und Schwächen für Ihren Anwendungsfall.

Hardware/Software-agnostisch bleiben: Es geht hier explizit nicht um spezifische Hard- oder Software. Sie bleiben auf der "Datenbank", "Server", etc. Ebene. Welche Datenbank (MySQL, PostgreSQL, SQLite, etc.) oder welcher Server (Apache, Node.js, etc.) verwendet wird kommt in der Realisierung. Auch hier: Nach der Konzeption kann der Leser stoppen und Ihre Konzeption auf seine Weise realisieren.

Bezug zum Interface Design beachten: Systemarchitektur und Interface Design hängen eng zusammen. Stimmen Sie beides aufeinander ab — in jeder Iteration.

Text ...

3.5 Interface Design

Wie können die Informationen dargeboten und Angaben der Nutzer erfolgen?

Screen-Fetisch hinterfragen: Viele Medieninformatiker haben einen Fetisch was visuelle Interface betrifft. Diese lassen sich leicht skizzieren, es gibt Tools dafür, die man kennt, und man kann was zeigen. Man sieht, was man gemacht hat. Für viele Kontexte sind sie allerdings absolut ungeeignet (z. B. um Autofahrer zu energiesparendem Fahren zu bringen). Visuelle Aufmerksamkeit ist extrem kostbar. Probieren Sie deswegen bewusst andere Sinne aus. Können Sie die Informationen auch auditiv geben (z. B. Abweichung vom idealen Fahrverhalten über die Tonhöhe)? Oder haptisch (z. B. Vibration)? Gerade wenn man sich im Laufe der Konzeption in das eigene visuelle Design verliebt kommen oft Lösungen heraus, die vielleicht schön sind, aber auch lebensgefährlich, und schlimmer noch: nicht gebrauchstauglich. Für viele Kontexte sind visuelle Interfaces sehr gut geeignet — aber für viele auch nicht. Und die Auswahl sollte wie üblich begründet erfolgen.

Formativ Evaluieren: Interface-Designs müssen Sie formativ evaluieren. Für Sie ist die Interaktion klar (Sie wissen auch als — meist — einzige Person, was unter der Oberfläche passiert und was die Hintergründe sind), für Laien ist das oft unklar. Sie brauchen früh Feedback der Zielgruppe. Nutzen Sie es! Dokumentieren Sie am Ende jeder Iteration, was herausgekommen ist und was die Konsequenzen für Ihr Design / die nächste Iteration war.

Papier über Mock-Up-Tools: Verwenden Sie Papier um visuelle Interfaces zu gestalten. Wenn Sie nicht zeichnen können, um so besser. Bei mit dem Computer ausgearbeiteten Mockups ist die Hemmung der Zielgruppe oft zu hoch, noch grundlegende Aspekte zu ändern. Damit bleiben Sie oft in der ersten — und oft unkreativen und nicht gebrauchstauglichen — Lösung gefangen. Tools wie Adobe XD sind eine Falle, die Sie zumindest in den ersten Iterationsphasen tunlichst vermeiden sollten.

Bei visuellen Interfaces Skizzen der Iterationen darstellen: Benutzen Sie Graphiken (Skizzen, Mockups) um das Interface und die Interaktion damit zu verdeutlichen.

Text ...

Nach der ersten Iteration: STOP

Nachdem Sie die erste Iteration abgeschlossen haben, nehmen Sie das Blatt Papier, auf dem Sie vor/während der Analyse Ihre Lösungsidee(n) festgehalten haben, wieder heraus. Vergleichen Sie Ihre ursprüngliche Idee mit der ersten Iteration.

Falls sich Ihre erste Iteration sich stark von der ersten Idee unterscheidet:

Prima. Sie haben während der Analyse dazu gelernt und den Nutzungskontext besser verstanden. Ohne die Analyse und die Verwendung dieser Informationen in der Konzeption hätten Sie das nicht machen können.

Falls Ihre erste Iteration Ihrer ersten Idee sehr ähnlich ist: Entweder Sie kannten sich schon vor der Analyse mit dem Themengebiet sehr gut aus — oder (und das ist wahrscheinlicher) — sie hängen zu stark an Ihrer ersten Idee und hören der Zielgruppe nicht zu. Auch falls Sie selbst zur Zielgruppe gehören (z. B. eine Anwendung für Studierende), Sie sind nicht "alle Nutzer". Nutzer unterscheiden sich üblicherweise. Entsprechend wirklich am in sich gehen — haben Sie aus der Analyse wirklich nichts dazu gelernt? Was das Problem (oder Sie selbst) wirklich so trivial? Das Schöne an kreativen Prozessen ist ja, dass man iterativ besser werden kann und auch auf Ideen kommen kann, auf die man vorher nicht gekommen wäre. Die Chance sollten Sie nutzen.

BTW, dieser Hinweis heißt nicht, dass Sie komplett naiv an Ihre Gestaltung herangehen sollten, ohne sich vorher Gedanken zu machen. Aber, wenn Sie eine Analyse wirklich ernst nehmen, dann sollte die Ihnen Sachverhalte aufzeigen, die man als Außen-stehender nicht sieht. Und Sie sollten auch in der Konzeption basierend auf diesen Informationen Lösungen entwickeln, die über die ersten simplen Ideen hinausgehen (vgl. confirmation bias). Die neu und nützlich sind.

3.6 Conclusion on the Conception

Wie auch bei der Analyse am Ende der Konzeption ein kurzes Fazit um den Entwicklungsprozess zusammenzufassen und begründen, was die erfolgversprechendste Lösung (höchste Gebrauchstauglichkeit in der formativen Evaluation!) für die Realisierung ist.

Text ...

4 Implementation

Beschreiben Sie, wie Sie das Projekt konkret realisiert haben, mit dem Ziel, anderen (Medien-)Informatikern zu erlauben, Ihre Anwendung weiter zu entwickeln.

Test: Sie nehmen einen Kommilitonen, geben ihm dieses Kapitel und bitten ihn eine Funktion hinzuzufügen oder zu verändern. Wenn er das kann, prima. Wenn nicht, besser schreiben oder bessere Kommilitonen suchen.

Verweisen Sie zu Beginn auf die erfolgreichste Lösung aus der Konzeption. Wie in der Konzeption bieten sich Unterkapitel mit Systemarchitektur und Interface-Design an — jetzt allerdings mit konkreten Angaben, welche Programme / Frameworks Sie verwenden, und wie Sie das Interface konkret umsetzen.

Geben Sie zuerst einen allgemeinen Überblick und werden dann konkreter. Als wenn Sie in eine Weltkarte reinzoomen. Erst mal sagen, wie die Welt aufgebaut ist (Kontinente), dann auf Kontinent mit Ländern zoomen, dann exemplarisch mal auf ein Land. Leser braucht ein gutes mentales Modell vom Aufbau und den Funktionen, um zu verstehen, wie es technisch umgesetzt wurde.

Begründen Sie Ihre Entscheidungen linear: Z.B.

1. "das sind die Anforderungen aus der Analyse/bisheriger Konzeption"
2. "Programmiersprache / System / etc. x, y, und z erfüllen diese Anforderungen"
3. "relevante Kriterien sind a, b, c und d (z. B. Verbreitung, Community, Weiterentwicklung, Preis)"
4. "Programmiersprache / System / etc. x, y, und z schneiden da so und so ab, x ist insgesamt am besten"
5. "Programmiersprache / System / SDK / etc. x wird verwendet". Eine Argumen-

tation à la $A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E$ ist nachvollziehbarer als wir nehmen E, weil $A \Rightarrow C$ woraus D folgt, wegen B.

Weiterer Grund: Wenn Sie zuerst die konkrete Software sagen, dann fangen Personen direkt an, die Entscheidung zu hinterfragen (weil sie z. B. eine andere bevorzugen, gibt da genug Religionskriege). Wenn Sie linear argumentieren gehen diese Personen eher bei jedem Schritt mit und enden dann bei der von Ihnen verwendeten Software (sofern sie bei den Auswahlkriterien und der Bewertung mitgehen).

Nutzen Sie den Datenträger im Anhang der Arbeit: Insbesondere in den Programmcode können Sie viel auslagern. In der Arbeit muss der Leser aber die Struktur/den Aufbau Ihres Programms verstehen sowie die Logik dahinter. Und er muss wissen, dass es die Erläuterungen im Programmcode gibt und wo er diese findet. Entsprechend die Struktur (Dateien/Klassen) deutlichen machen und Verweise auf relevanten Dateien.

Text ...

4.1 System Architecture Implementation

vgl. Konzeption, jetzt aber wie und womit es konkret umgesetzt wurde (u.U. bieten sich UML-Diagramme an)

Text ...

4.2 Interface Implementation

vgl. Konzeption, jetzt aber wie und womit es konkret umgesetzt wurde

Text ...

4.3 Conclusion on the Implementation

Kurzes Fazit, das u.a. die Frage beantwortet, was technisch konkret umgesetzt wurde und was z. B. aus der Konzeption übrig geblieben ist und warum.

Text ...

5 Dialogue Samples

Verdeutlichen Sie die Funktionsweise Ihrer Entwicklung bei einer typischen Bedienung (z. B. wie in einem Anwendungsbeispiel / Use Case beschrieben — auf den können Sie sich ja beziehen).

Die Abbildungen müssen im Text vorher erläutert sein und einen Eindruck geben, wie die Entwicklung konkret bedient wird bzw. was sie kann.

Das ist das, was bleibt: Diese Dialogbeispiele sind das einzige, was in ein paar Jahren von Ihrer Anwendung noch sichtbar ist. Dann ist entweder der Datenträger verschwunden oder beschädigt oder die notwendige Hard- und Software ist nicht mehr lauffähig. Geben Sie dem Leser entsprechend einen guten Einblick in das, was Sie tatsächlich realisiert haben.

Jetzt ein Video aufnehmen: Schalten Sie den Screenrecorder auf dem Computer (not-falls: QuickTime) oder Smartphone (kann das OS) an und nehmen Sie die typischen Interaktionen einmal auf. Das ist Ihr "Plan B", falls (oft: wenn) im Kolloquium die App nicht bedient werden kann. Das können Sie auch gut auf DVD brennen um Lesern die Möglichkeit geben, sich die Interaktion anzusehen (in der Arbeit darauf hinweisen!). Macht sich auch gut auf Websites (insbesondere dem eigenen Portfolio).

Text ...

6 Summative Evaluation

Geben Sie zu Beginn der Evaluation einen kurzen Überblick über Ihr Vorgehen. Dazu reichen meist die Zwischenüberschriften mit ein oder zwei Sätzen, was Sie konkret gemacht haben. Also nicht "In Design wird das Design beschrieben" sondern "Das experimentelle Vorgehen wird im Abschnitt Design dargestellt".

Text ...

6.1 Goal

Die summative Evaluation ist eine abschließende Bewertung Ihrer Entwicklung. Ziel ist, unter dem Strich zu sehen, wie gebrauchstauglich Ihr System ist (nicht mehr eine iterative Verbesserung wie in der formativen Evaluation der Konzeption). Da-für müssen Sie klare Kriterien ableiten, was eine "gute" bzw. "schlechte" Bewertung nach sich ziehen würde. Meist sind das die klassischen Gebrauchstauglichkeitskriterien (Effektivität, Effizienz, Erlernbarkeit, Zufriedenstellung), wobei diese über das Ziel Ihrer Anwendung konkretisiert werden (Effektivität bei einer Lernapp ist konkret gemessen etwas anderes als Effektivität bei einem digitalen Depressionstagebuch).

Hier verdeutlichen Sie entsprechend, welche Fragen die Evaluation beantworten soll.

Text ...

6.2 Methods

Im Methodenteil zeigen Sie, was wie evaluiert wurde. Der Methodenteil muss anderen Entwicklern die Möglichkeit geben, Ihr Evaluationsvorgehen zu wiederholen um Ihre Ergebnisse zu überprüfen.

Text ...

6.2.1 Design

Kurze Beschreibung des Versuchs- oder Evaluationsdesigns, dass man das Vorgehen einordnen kann. Also z. B. "es wurde ein Usability Test durchgeführt", oder "es wurde ein Experiment mit between-subjects design durchgeführt, bei dem die Kontrollgruppe die bisherige App verwendet hat und die Experimentalgruppe die neu entwickelte App".

Text ...

6.2.2 Participants

Kurze Beschreibung der Teilnehmer mit relevanten Angaben. In jedem Fall die Anzahl, oft noch Geschlecht, Alter, Beruf, Vorerfahrung, etc. Bei weniger als zehn Teilnehmern bietet sich eine Tabelle zur schnellen Übersicht an. Verweisen Sie bei individuellen Ergebnissen (z.B. Zitaten aus Fragebögen oder Interviews) auf die Teilnehmer-Nummer.

Achtung: Niemals die Namen der Teilnehmer erwähnen! Die Teilnehmer stehen stellvertretend für die Zielgruppe. Wer sie konkret sind ist irrelevant. Begründen Sie, warum Sie diese Personen ausgewählt haben (spiegeln die Nutzergruppe gut wider) und wo/wie Sie diese rekrutiert haben.

Text ...

6.2.3 Setting and Instruments

Beschreiben Sie die notwendigen Materialien bei der Evaluation. Dazu gehört — mit Überschriften klar ausgewiesen — das Setting (wo wurde die Evaluation durchgeführt),

Ihre Entwicklung (Verweis auf Dialogbeispiele), und Ihre Erhebungsmethoden (Fragebögen, Interviewleitfäden, etc.).

Falls Sie etablierte Fragebögen verwenden (z.B. ATI) reicht die entsprechende Zitation mit einer kurzen Beschreibung. Bei längeren Fragebögen oder Interviewleitfäden nennen Sie kurz die Abschnitte (z.B. soziodemographische Daten, Technikerfahrung, etc.) und verweisen Sie auf die vollständigen Fragebögen im Anhang.

ACHTUNG: Zeitliche Reihenfolge ist hier egal. Hier geht es nach Gliederungspunkten wie Instrumente (Fragebögen, Interviews, etc.). Die zeitliche Reihenfolge wird in der Prozedur dargestellt.

Setting

Text ...

Verwendete Anwendung

Text ...

Fragebögen

Text ...

6.2.4 Procedure

Beschreiben Sie chronologisch den Ablauf der Evaluation, von der Begrüßung bis zur Verabschiedung. Verweisen Sie dabei auf die anderen Abschnitte (v.a. Setting und Instrumente) und führen Sie nichts Neues mehr ein. Dieser Abschnitt ist der einzige Abschnitt, in dem Sie die zeitliche Reihenfolge klar einhalten müssen, alle anderen sind inhaltlich strukturiert.

Text ...

Nach dem Lesen der Methode muss deutlich geworden sein, wie Sie Ihre Evaluationsfragen messbar gemacht haben. Wie haben Sie z.B. Benutzerzufriedenheit oder Effizienz

gemessen? Die Leser müssen sich aufgrund des Methodenteils in die Lage der Teilnehmer versetzen können und ein mentales Modell Ihrer Evaluation bilden können.

6.3 Results

In den Ergebnissen zeigen Sie wie die Ergebnisse analysiert wurden um Ihre Evaluationsfragen zu beantworten. Nicht einfach die Daten auflisten, sondern stellen Sie die Ergebnisse strukturiert dar und betonen Sie die wichtigen Aspekte. Gliedern Sie die Ergebnisse nach Ihren Forschungsfragen / Fragestellungen (nicht zeitlich oder nach Erhebungsmethoden wie Fragebögen vs. Beobachtung). Zu Beginn (falls relevant) sollten Sie überprüfen, ob die Anwendung auch wirklich so verwendet wurde, wie sie verwendet werden sollte (manipulation check). Haben die Personen also z. B. wirklich die Aufgaben mit der App gelöst oder haben sie die App schnell beiseite gelegt.

Stellen Sie die Befunde / Ergebnisse von Analysen / Evaluationen / etc. immer so neutral und so objektiv wie möglich dar — ohne Ihre subjektive Interpretation oder Bewertung. Also keine Begriffe wie "lediglich", "nur", "könnte / würde / sollte / etc." oder Bewertungen wie "hat gut / nicht gut geklappt". Diese Bewertungen gehört in die Diskussion.

Überlegen Sie sich mit welchen Tabellen und Abbildungen Sie die Ergebnisse gut darstellen können. Bei aggregierten Messwerten (auf mindestens Intervallskalenniveau) immer Mittelwerte (M), Standardabweichungen (SD) und die Anzahl der Messdaten (Personen, n) angeben. Bei ordinalskalierten Daten entsprechend Median, etc. (ja, Statistik war wichtig).

Statistische Tests korrekt angeben (siehe z.B. Pallant, 2010).

Pallant, J. (2007). SPSS Survival Manual (3rd ed.). Open University Press.

Abbildungen müssen in sich verständlich sein (was abgebildet ist). Das heißt, die Achsen eindeutig beschriften, Skala (z.B. Likert-Skala von 1 starke Ablehnung bis 7 starke Zustimmung) in die Legende. 3D-Graphiken vermeiden — diese bieten oft keinen Mehrwert (vgl. Field, 2016).

Field, A. (2016). An Adventure in Statistics. Sage.

Am Ende des Ergebnis-Abschnitts muss deutlich geworden sein, wie Ihre Entwicklung von den Teilnehmenden eingesetzt wurde (hoffentlich wie geplant), was die Hauptergebnisse waren, sofern aufgrund der Stichprobengröße möglich welche Werte sich statistisch signifikant voneinander unterscheiden und was die statistischen Ergebnisse in den Variablen bedeuten (z. B. positive Korrelation zwischen A und B, dass Personen, die A besser bewertet haben auch B besser bewertet haben; aber keine Bewertung ob das gut oder schlecht ist). Falls Sie konkrete Ziele hatten (z. B. "SUS-Wert von x") dann sagen Sie, ob dieses Ziel erreicht wurde oder nicht (das ist keine Wertung, die in die Diskussion gehören würde, sondern ein größer, gleich oder kleiner was eindeutig ist).

Text ...

6.4 Discussion

In der Diskussion erklären und interpretieren Sie die Ergebnisse. Welche Schlussfolgerungen ziehen Sie daraus? Was sind die praktischen Konsequenzen für die (weitere) Entwicklung? Hier dürfen Sie selbst die Ergebnisse bewerten — auf Basis von einer kritischen Reflektion. In der Diskussion keine neuen Ergebnisse aus der Analyse / Evaluation / etc. einführen. Die Beweisaufnahme ist mit Ende des Ergebnisteils abgeschlossen. Es geht hier auch nicht um eine Mystery-Geschichte mit Spannungsbogen, sondern um klar nachvollziehbare Argumente. Neue Informationen aus der Literatur verwenden um die (v.a. überraschende) Ergebnisse zu interpretieren ist dagegen möglich.

Text ...

6.5 Conclusion on the Evaluation

Fassen Sie die Evaluation kurz zusammen — insbesondere was die zentralen Ergebnisse waren. Unterm Strich: Wie gut hat's geklappt?

Text ...

7 Summary and Outlook

Kurze Einführung, was in den folgenden Unterkapiteln behandelt wird.

Text ...

7.1 Summary

Fassen Sie die zentralen Schritte und Ergebnisse Ihrer Arbeit kurz zusammen. Personen mit wenig Zeit müssen aus dieser Darstellung die Kernpunkte Ihrer Arbeit mitnehmen und Ihren Arbeitsaufwand und Erfolg bewerten können.

Ist ähnlich wie die Zusammenfassung zu Beginn der Arbeit, aber etwas länger (1 bis maximal 2 Seiten) mit Verweisen auf die entsprechenden Kapitel/Abschnitte und Sie können beim Leser etwas mehr voraussetzen (hat es gelesen oder kann wegen den Verweisen direkt dahin springen).

Text ...

7.2 Outstanding Issues

Offene Punkte = Versprochene aber nicht umgesetzte Punkte: Falls Schritte explizit geplant wurden (Exposé! Pflichtenheft!), aber nicht realisiert werden konnten, dann diese hier klar darstellen und diskutieren.

Mögliche Features, die Sie nicht vor Beginn versprochen haben, gehören in den Ausblick.

Text ...

7.3 Outlook

Ideen, welche weiteren Entwicklungen oder Untersuchungen folgen sollten, oder was man noch umsetzen könnte, gehören in den Ausblick.

Versprochene aber nicht umgesetzte Elemente in die Offenen Punkte.

Bitte keine Allgemeinheiten ("können noch Features hinzugefügt werden" oder "könnte besser evaluiert werden") sondern konkrete Beschreibungen und Begründungen der Relevanz dieser Schritte.

Text ...

7.4 Final Conclusion

Die Arbeit, die Ergebnisse und weitere mögliche Schritte kurz kritisch (ehrlich und konstruktiv, aber nicht selbstkreuzigend) reflektieren und positiv enden. Maximal eine halbe bis 3/4 Seite. Ist kurz und hier können Sie von der Arbeit zurücktreten und auch den Leser aus dem Text ziehen.

Das ist nur zum Test: Ford (2011) Apache Software Foundation (2010)

Text ...

List of Figures

1.1	AI Control Room - Model Assessment Overview with Standard Metrics .	5
1.2	AI Control Room - Precision-Recall and Calibration Graphs	6
1.3	AI Control Room - Model String Points and Limitations	6
1.4	AI Control Room - Interpretability Assessment	7
1.5	AI Control Room - Example Data	7
1.6	AI Control Room - Prediction Explanation for Examples Data	7
1.7	Human-centered Design Process (<i>DIN EN ISO 9241-210</i> , 2011)	9

List of Tables

Sources

References

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Alley, M. (1996). *The Craft of Scientific Writing* (Vol. 3). Springer.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *arXiv:1910.10045 [cs]* arxiv 1910.10045. Retrieved May 19, 2021, from <http://arxiv.org/abs/1910.10045>
- Clearbox AI. (2021a). *Clearbox AI Model Assessment* (Whitepaper). Retrieved April 2021, from https://clearbox.ai/pdf/ClearboxAI_Technical_Whitepaper.pdf
- DIN EN ISO 9241-210: *Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*. (2011). DIN Deutsches Institut für Normung e. V. Beuth Verlag.
- European Commission. (2020). *On Artificial Intelligence - A European approach to excellence and trust* (Whitepaper). Retrieved September 2021, from https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Gordon, M. L., Zhou, K., Patel, K., Hashimoto, T., & Bernstein, M. S. (2021, May 6). The disagreement deconvolution: Bringing machine learning performance metrics in line with reality, In *Proceedings of the 2021 CHI conference on human factors in computing systems*. CHI '21: CHI Conference on Human Factors in Computing Systems, Yokohama Japan, ACM. <https://doi.org/10.1145/3411764.3445423>
- Greenberg, S., Carpendale, S., Marquardt, N., & Buxton, B. (2012). *Sketching User Experiences. The Workbook*. Morgan Kaufmann.

- Keane, M. T., & Kenny, E. M. (2019). How case based reasoning explained neural networks: An XAI survey of post-hoc explanation-by-example in ANN-CBR twins. *arXiv:1905.07186 [cs]*, 11680arxiv 1905.07186, 155–171. https://doi.org/10.1007/978-3-030-29249-2_11
- Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable artificial intelligence for human decision-support system in medical domain. *arXiv:2105.02357 [cs]*arxiv 2105.02357. Retrieved June 11, 2021, from <http://arxiv.org/abs/2105.02357>
- Knapp, J., Zeratsky, J., & Kowitz, B. (2016). *Sprint. How to solve big problems and test new ideas in just five days*. Simon & Schuster Paperbacks.
- Kumar, V. (2013). *101 Design Methods - A Structured Approach for Driving Innovation in Your Organization*. John Wiley & Sons, Inc.
- Lapuschkin, S., Binder, A., Montavon, G., Müller, K.-R., & Samek, W. (2016). The lrp toolbox for artificial neural networks. *Journal of Machine Learning Research*, 17(114), 1–5. <http://jmlr.org/papers/v17/15-618.html>
- Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., & Müller, K.-R. (2019). Unmasking clever hans predictors and assessing what machines really learn. *Nature Communications*, 10(1), 1096. <https://doi.org/10.1038/s41467-019-08987-4>
- Ras, G., van Gerven, M., & Haselager, P. (2018). Explanation methods in deep learning: Users, values, concerns and challenges. *arXiv:1803.07517 [cs, stat]*arxiv 1803.07517. Retrieved May 19, 2021, from <http://arxiv.org/abs/1803.07517>
- Ras, G., Xie, N., van Gerven, M., & Doran, D. (2021). Explainable deep learning: A field guide for the uninitiated. *arXiv:2004.14545 [cs, stat]*arxiv 2004.14545. Retrieved October 6, 2021, from <http://arxiv.org/abs/2004.14545>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. *arXiv:1602.04938 [cs, stat]*arxiv 1602.04938. Retrieved May 19, 2021, from <http://arxiv.org/abs/1602.04938>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1). Retrieved June 24, 2021, from <https://ojs.aaai.org/index.php/AAAI/article/view/11491>
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278. <https://doi.org/10.1109/JPROC.2021.3060483>

Websites

- Clearbox AI. (2021b, October). *Manage AI models with confidence*. <https://clearbox.ai/>
- CoCoAI. (2021, September). *Cooperative and communicating AI methods for medical image-guided diagnostics - A research project at the University of Lübeck*. <https://cocoai.uni-luebeck.de>
- European Commision. (2021, March). *Ethics guidelines for trustworthy AI*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- Ford, R. (2011, August). *Earthquake: Twitter Users Learned of Tremors Seconds Before Feeling Them*. <http://www.hollywoodreporter.com/news/earthquake-twitter-users-learned-tremors-226481>
- ThinkSono. (2021, September). *The world's first software to detect DVT!* <https://thinksono.com/>

Software

- Apache Software Foundation. (2010, February 19). *Hadoop* (Version 0.20.2). <https://hadoop.apache.org>

Software-Name mit Versionsnummer und Link zur Website. Nur was für die konkrete Arbeit relevant ist. Das Sie die Arbeit mit Word geschrieben haben, ist irrelevant. Das sieht man. LaTeX auch.

Abbreviations

AI Artificial Intelligence

AUROC Area Under the Receiver Operating Characteristics

CNN Convolutional Neural Network

DNN Deep Neural Network

DVT Deep Vein Thrombosis

ML Machine Learning

XAI Explainable Artificial Intelligence

Glossary

Artificial Intelligence Systems that display intelligent behaviour by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

Machine Learning Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

Area Under the Receiver Operating Characteristics Important evaluation metric for checking any classification model's performance.

Black Box System which can be viewed in terms of its inputs and outputs (or transfer characteristics), without any knowledge of its internal workings.

Convolutional Neural Network Convolutional neural networks are a class of artificial neural networks, most commonly used for processing structured arrays of data such as images..

Deep Neural Network Deep neural networks are a powerful category of machine learning algorithms implemented by stacking layers of neural networks along the depth and width of smaller architectures.

Deep Vein Thrombosis Deep vein thrombosis occurs when a blood clot forms in one or more of the deep veins in your body, usually in the legs.

Explainable Artificial Intelligence Artificial intelligence in which the results of the solution can be understood by humans.

Appendices

Zusätzliche Informationen die zu lang für die Arbeit sind können hier verfügbar gemacht werden.

Aber auch an die DVD denken — was ist dort besser aufgehoben? Die Zeiten, in denen man Programmcode manuell eingetippt hat, sind ja glücklicherweise lange vorbei, deswegen macht Code hier wenig Sinn.

Ist entsprechend ein Priorisierung: Was würde sich der Leser vielleicht gerne während des Lesens der Arbeit (z. B. im Zug) ansehen, wenn er auch gerade nicht auf die DVD zugreifen kann (kein DVD Laufwerk)?

Inhalte sind oft: Überblick der Inhalte der DVD, Fragebögen (falls digital Screenshots oder neu für den Druck formatiert), Interviewleitfäden, etc. Selten detailliertere Evaluationsergebnisse.

Hier kurz die Zwischenüberschriften nennen und evtl. 1 Satz, was dort zu finden ist (falls es nicht schon durch die Zwischenüberschrift klar ist).

Text ...

Appendix A: Inhalt der DVD

Oft ein Default: Was findet man auf der beiliegenden DVD in welchem Verzeichnis? Max. 1 Seite.

In jedem Fall die PDF der Arbeit, den Programmcode, Daten (anonymisiert!).

Niemals Interviewaufzeichnungen, Einverständniserklärungen oder ähnliche personen-

bezogene Daten auf die DVD brennen — Sie haben in den meisten Fällen Anonymität zugesichert und die DVD ist frei zugänglich (ein Exemplar der Arbeit kommt in die Bibliothek).

Text ...

Appendix B: Anhangstitel

Weitere Inhalte je nachdem, wo der Leser ohne großen Aufwand hinspringen sollte.

Text ...

Assertion under Oath

I declare in lieu of an oath that I have written this paper independently and have used only the sources indicated.

[Nach Ausdruck unterschreiben. Muss auf Papier sein.]

Lübeck, 6th October, 2021, Philipp Dominik Bzdok