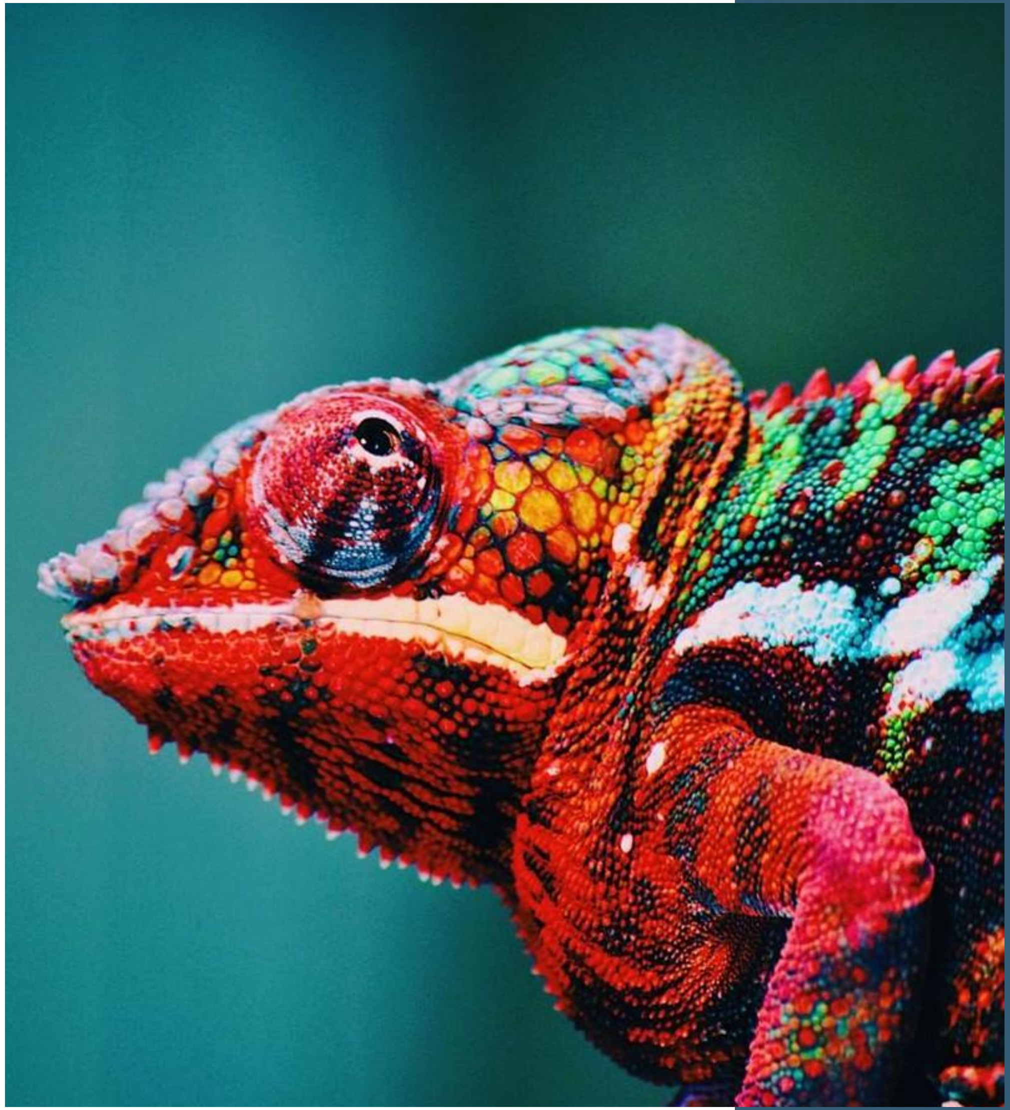
Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal

Frank Binder[1], Jana Diels[2]
[1]Institute for Applied Informatics (InfAI) at Leipzig University, Germany
[2]ConPolicy GmbH – Institute for Consumer Policy, Berlin, Germany

Presentation at “AI-in-the-loop - Reconfiguring HCI for AI development”, Panel at “Culture and Computing”. Part of HCII 2022.



Contacts and acknowledgements

Frank Binder
Speaker

Research Associate

InfAI®
Institut für Angewandte Informatik

Dr. Jana Diels
Attending

Project Manager

ConPolicy
Institut für Verbraucherpolitik

Joint work with
Julian Balling[1,3], Oliver Albrecht[4], Robert Sachunsky[3], J. Nathanael Philipp[3], Yvonne Scheurer[4], Marlene Münsch[2], Markus Otto[5], Andreas Niekler[3], Gerhard Heyer[1,3], and Christian Thorun[2]

[1] Institute for Applied Informatics at Leipzig University (InfAI), Germany
[2] ConPolicy GmbH - Institute for Consumer Policy, Berlin, Germany
[3] Natural Language Processing Group, Leipzig University, Germany
[4] co2online gGmbH, Berlin, Germany
[5] SEnerCon GmbH, Berlin, Germany

Gefördert durch:

Bundesministerium
der Justiz und
für Verbraucherschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Plan

1. Motivation
2. Definitions: Machine Learning-Based Systems and “the Loop”
3. Case study and methods
4. Applying User Research to Guide the ML Development Process
5. Measuring User Success and ML Performance
6. Limitations, Discussion and Outlook

Putting Users in the Loop

Motivation

- Evaluation of ML-based systems ususally focuses on accuracy (and speed) of processing
- User perspective is often neglected or apparently difficult to capture, esp. for SMEs and Public Sector
- Academic paradigms of ML evaluation provide little assistance in tackling the „data sourcing dilemma“ for real-life ML-based services
- Fuzziness in public discussion about the nature of „AI“ results in premature conclusions or inappropriate expectations (even among stakeholders)
- Unique challenges for different “AI projects”

→ Our approach and contribution

- Case study on how to include user research in development process of ML-based systems
- Incremental approach to keep users involved and collect data even while the ML-component is of limited use
- ML quality and user acceptance are mutually interdependent
- Putting users in the loop becomes mandatory!
- Measuring user success and ML performance allows to dynamically adjust workload
- Check expectations and generalizations, e.g. by applying a typology of ML-enabled use cases

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal.
Culture and Computing. HCII 2022.

ML-based Systems

‘The Loop’

Roles within the loop

Use case / case study

User research

Measure user success and ML performance

Machine Learning (ML)-based Systems

- As defined by Riccio et al. 2021, ref. [29]:

  ML-based systems are software systems that include “one or more [software] components that learn how to perform a task from a given data set”

- Similarly:

  ML-based service denotes a digitally provided application or service that assists users in reaching a particular goal by using the respective ML-based system

- Zooming in:

  ML component is the respective ‚ML part‘ of the ML-based system

- Even further:

  ML model (within ML component) captures the aquired knowledge

ML-based Systems

‘The Loop’

Roles within the loop

Use case / case study

User research

Measure user success
and ML performance

Concepts & methods

What is ‘the Loop’?

ML components are developed and maintained
in an **iterative process**, which we call ***the Loop***:

1. Collect and curate a new/revised set of training data
2. Take an ML model - possibly from the production setting - and re-train or fine-tune it on this newly compiled data set
3. Evaluate the model’s performance with regard to accuracy, speed (or computing resources used, etc.)
4. In case of improvements, re-deploy the newly created ML model increment to the production setting
5. Start over

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

ML-based Systems

‘The Loop’

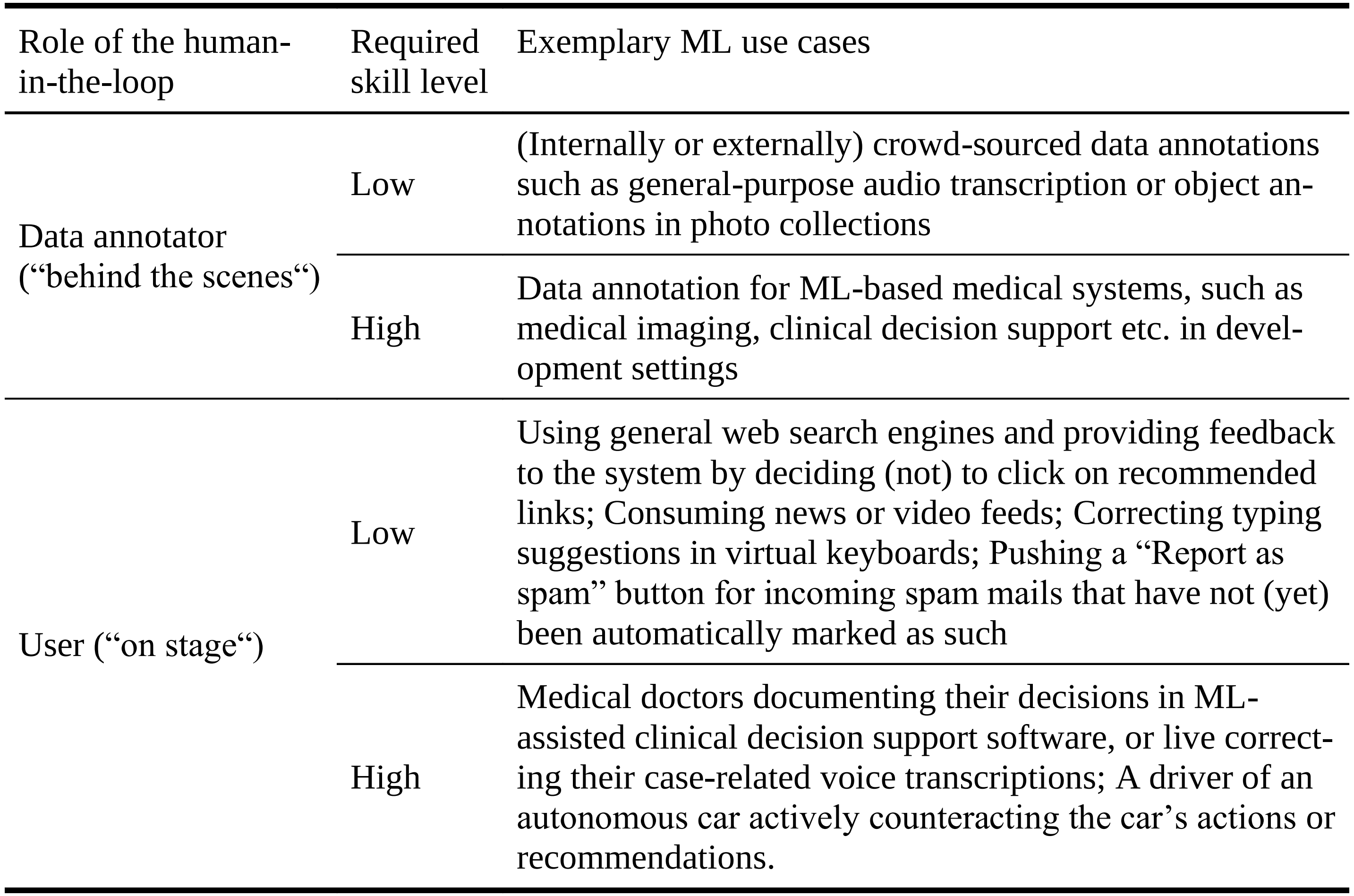
Roles within the loop

Use case / case study

User research

Measure user success and ML performance

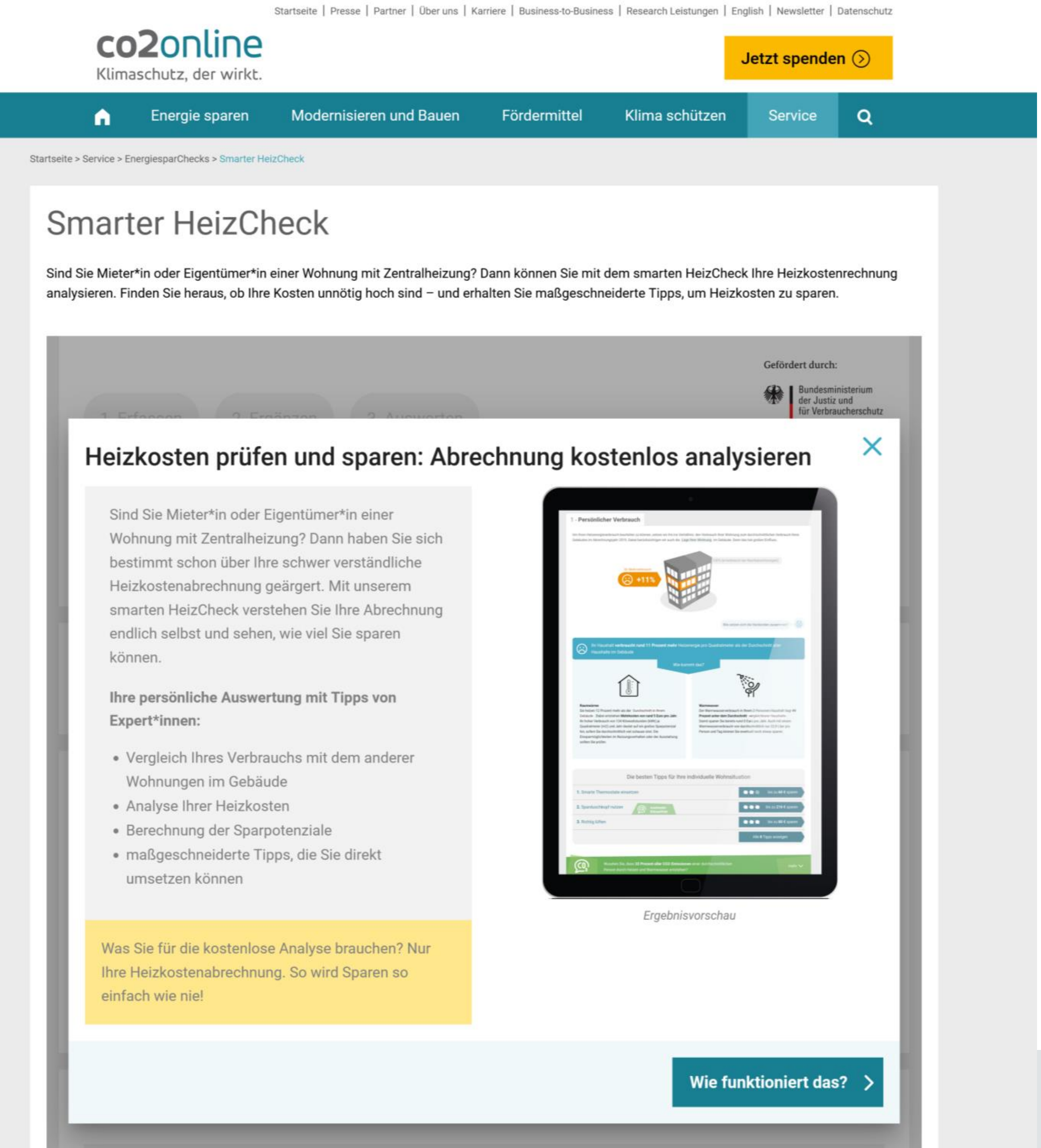
Where is the human in the Loop?



Role of the human-in-the-loop	Required skill level	Exemplary ML use cases
Data annotator (“behind the scenes“)	Low	(Internally or externally) crowd-sourced data annotations such as general-purpose audio transcription or object annotations in photo collections
	High	Data annotation for ML-based medical systems, such as medical imaging, clinical decision support etc. in development settings
User (“on stage“)	Low	Using general web search engines and providing feedback to the system by deciding (not) to click on recommended links; Consuming news or video feeds; Correcting typing suggestions in virtual keyboards; Pushing a “Report as spam” button for incoming spam mails that have not (yet) been automatically marked as such
	High	Medical doctors documenting their decisions in ML-assisted clinical decision support software, or live correcting their case-related voice transcriptions; A driver of an autonomous car actively counteracting the car’s actions or recommendations.

See references [6,14,15,21,35] in paper.

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*


Startseite | Presse | Partner | Über uns | Karriere | Business-to-Business | Research Leistungen | English | Newsletter | Datenschutz
co2online
Klimaschutz, der wirkt.
Jetzt spenden
Energie sparen
Modernisieren und Bauen
Fördermittel
Klima schützen
Service
Startseite > Service > EnergiesparChecks > Smarter HeizCheck
1. Erfassen
2. Ergänzen
3. Auswerten
Gefördert durch:
Bundesministerium der Justiz und für Verbraucherschutz
Ergebnisvorschau
Wie funktioniert das?

Smarter HeizCheck

Sind Sie Mieter*in oder Eigentümer*in einer Wohnung mit Zentralheizung? Dann können Sie mit dem smarten HeizCheck Ihre Heizkostenrechnung analysieren. Finden Sie heraus, ob Ihre Kosten unnötig hoch sind – und erhalten Sie maßgeschneiderte Tipps, um Heizkosten zu sparen.

Heizkosten prüfen und sparen: Abrechnung kostenlos analysieren

Sind Sie Mieter*in oder Eigentümer*in einer Wohnung mit Zentralheizung? Dann haben Sie sich bestimmt schon über Ihre schwer verständliche Heizkostenabrechnung geärgert. Mit unserem smarten HeizCheck verstehen Sie Ihre Abrechnung endlich selbst und sehen, wie viel Sie sparen können.

Ihre persönliche Auswertung mit Tipps von Expert*innen:

- Vergleich Ihres Verbrauchs mit dem anderer Wohnungen im Gebäude
- Analyse Ihrer Heizkosten
- Berechnung der Sparpotenziale
- maßgeschneiderte Tipps, die Sie direkt umsetzen können

Was Sie für die kostenlose Analyse brauchen? Nur Ihre Heizkostenabrechnung. So wird Sparen so einfach wie nie!

Smart_HEC

Case study:
Heating cost check

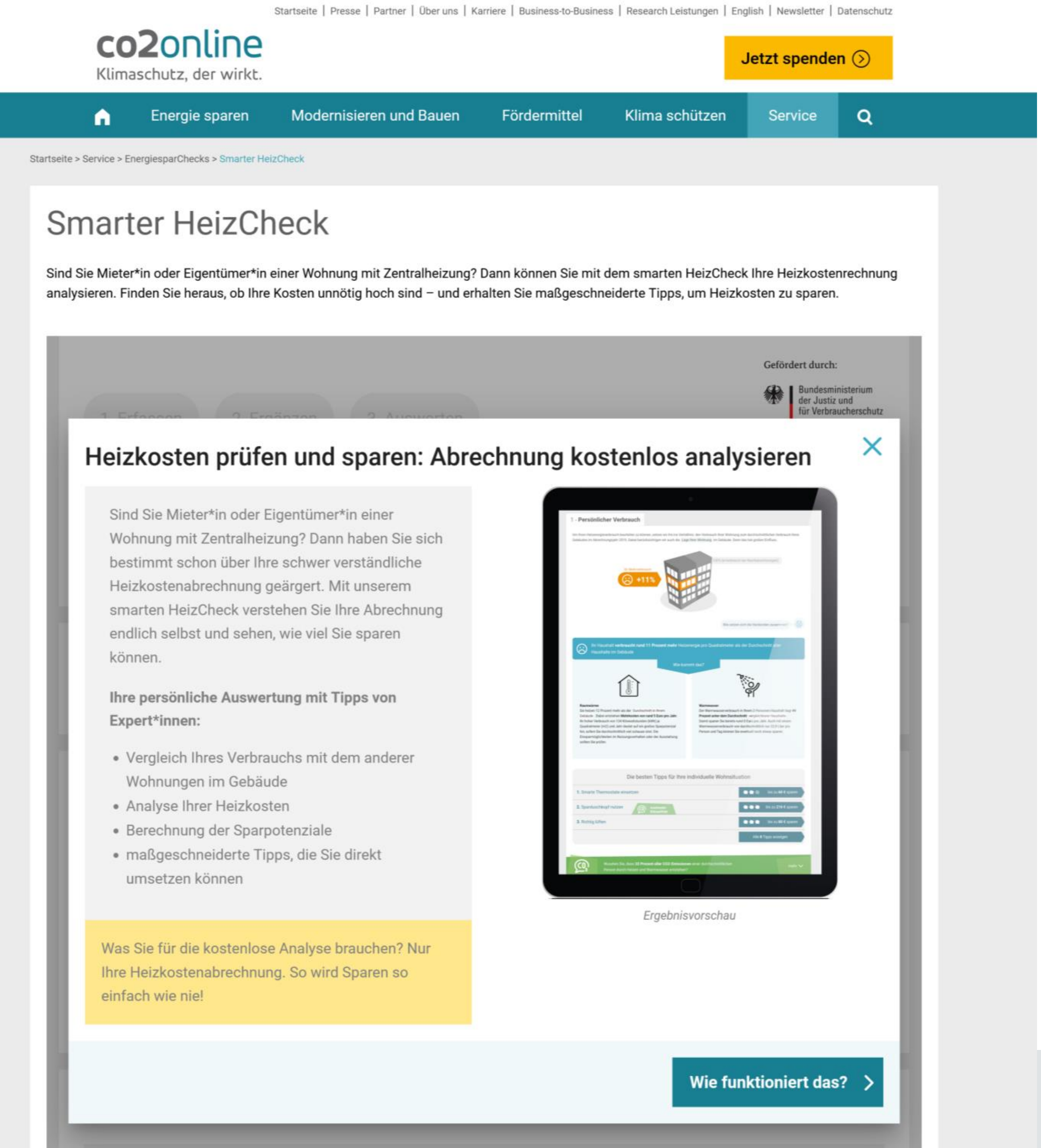
Consumers can analyze their heating bills for potential energy and cost savings

- Users upload their heating bills
- Specific data values are extracted*
- An assessment of potential savings is derived and reported back to the user
- A personalized recommended course of actions is generated

* Data extraction is performed by a highly customized open-source machine learning component for visual document analysis: OCR-D [9, 24]

https://www.co2online.de/service/energiesparchecks/smarter-heizcheck/

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*


Startseite | Presse | Partner | Über uns | Karriere | Business-to-Business | Research Leistungen | English | Newsletter | Datenschutz
co2online
Klimaschutz, der wirkt.
Jetzt spenden
Energie sparen
Modernisieren und Bauen
Fördermittel
Klima schützen
Service
Startseite > Service > EnergiesparChecks > Smarter HeizCheck
Gefördert durch:
Bundesministerium der Justiz und für Verbraucherschutz
1. Erfassen
2. Ergänzen
3. Auswerten
1 - Persönlicher Verbrauch
+11%
Ergebnisvorschau
Wie funktioniert das?

Smarter HeizCheck

Sind Sie Mieter*in oder Eigentümer*in einer Wohnung mit Zentralheizung? Dann können Sie mit dem smarten HeizCheck Ihre Heizkostenrechnung analysieren. Finden Sie heraus, ob Ihre Kosten unnötig hoch sind – und erhalten Sie maßgeschneiderte Tipps, um Heizkosten zu sparen.

Heizkosten prüfen und sparen: Abrechnung kostenlos analysieren

Sind Sie Mieter*in oder Eigentümer*in einer Wohnung mit Zentralheizung? Dann haben Sie sich bestimmt schon über Ihre schwer verständliche Heizkostenabrechnung geärgert. Mit unserem smarten HeizCheck verstehen Sie Ihre Abrechnung endlich selbst und sehen, wie viel Sie sparen können.

Ihre persönliche Auswertung mit Tipps von Expert*innen:

- Vergleich Ihres Verbrauchs mit dem anderer Wohnungen im Gebäude
- Analyse Ihrer Heizkosten
- Berechnung der Sparpotenziale
- maßgeschneiderte Tipps, die Sie direkt umsetzen können

Was Sie für die kostenlose Analyse brauchen? Nur Ihre Heizkostenabrechnung. So wird Sparen so einfach wie nie!

Smart_HEC

Case study:
Main Hypotheses

“In real market settings for consumer-oriented applications, AI quality and user acceptance are interlinked and need to be treated as mutually interdependent throughout the entire development process.”

https://www.co2online.de/service/energiesparchecks/smarter-heizcheck/

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

User Research: Goals & methods

References regarding the applied methods: [3,8,16,18,22,26]

Preparatory stage

(1) Outline individual needs & expectations

→ User Story Mapping Workshop

(2) Understand motivational levers

→ Explorative (online) focus groups

(3) Reveal concerns & expectations on data protection standards

→ Quant. online survey

Stage 1: Beta

(1) Realistically test user interaction (guidance, navigation & functionality)

→ UX Tests (Thinking Aloud approach)

(2) Reveal & verify added value

→ (Online) focus groups

Stage 2: Prototype

(1) Observe users live interaction

(2) Identify critical points with high prob. of abandonment and understand reasons

(3) Test digital-native vs. non-natives

→ UX Tests (Thinking Aloud approach)

Results feed back into techical development

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

User Research: Key findings

UX-test in stage 1 (early beta):

Users had difficulties with uploading or inserting data

- Data submission / document upload process was revised

Focus groups:

Test persons did not fully understand the results,
lost interest to further engage or act upon results

- Overhaul results section to be more relevant and actionable

UX-test in stage 2 (final prototype):

Both user groups (digital natives and non-natives) place
high demands on both ML accuracy and speed,
as well as on general functionality of the web app

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

ML-based Systems

‘The Loop’

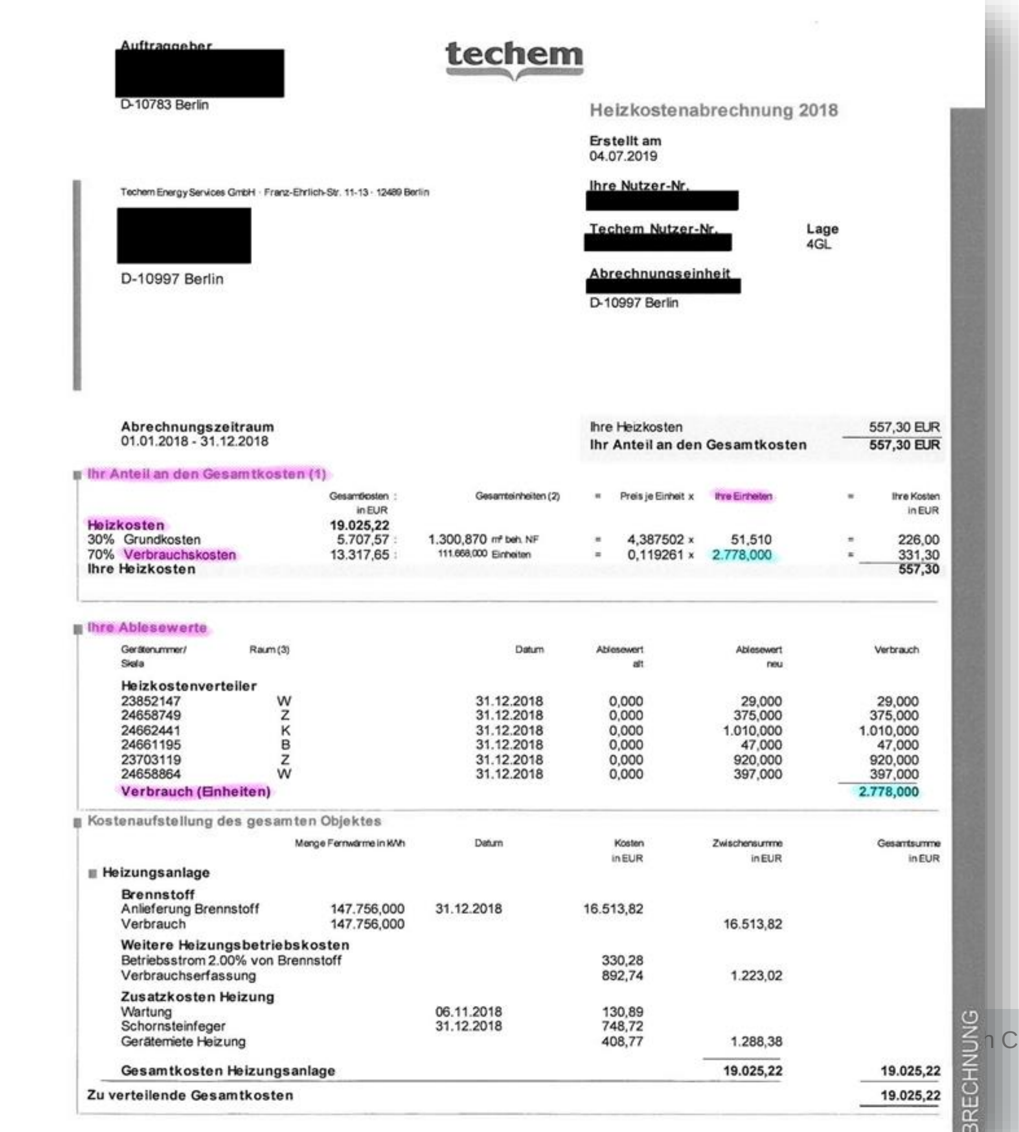
Roles within the

Use case / case

User research

Measure user su
and ML performa

Measuring user success and ML performance



Auftraggeber
[redacted]
D-10783 Berlin

techem

Heizkostenabrechnung 2018

Erstellt am
04.07.2019

Techem Energy Services GmbH · Franz-Ehrlich-Str. 11-13 · 12489 Berlin

[redacted]
D-10997 Berlin

Ihre Nutzer-Nr.
[redacted]

Techem Nutzer-Nr.
[redacted]

Lage
4GL

Abrechnungseinheit
[redacted]
D-10997 Berlin

Abrechnungszeitraum
01.01.2018 - 31.12.2018

Ihre Heizkosten 557,30 EUR
Ihr Anteil an den Gesamtkosten 557,30 EUR

Ihr Anteil an den Gesamtkosten (1)

	Gesamtkosten : in EUR	Gesamteinheiten (2)	=	Preis je Einheit x	Ihre Einheiten	=	Ihre Kosten in EUR
Heizkosten	19.025,22						
30% Grundkosten	5.707,57 :	1.300,870 m² beh. NF	=	4,387502 x	51,510	=	226,00
70% Verbrauchskosten	13.317,65 :	111.668,000 Einheiten	=	0,119261 x	2.778,000	=	331,30
Ihre Heizkosten							557,30

Ihre Ablesewerte

Gerätenummer/ Skala	Raum (3)	Datum	Ablesewert alt	Ablesewert neu	Verbrauch
Heizkostenverteiler					
23852147	W	31.12.2018	0,000	29,000	29,000
24658749	Z	31.12.2018	0,000	375,000	375,000
24662441	K	31.12.2018	0,000	1.010,000	1.010,000
24661195	B	31.12.2018	0,000	47,000	47,000
23703119	Z	31.12.2018	0,000	920,000	920,000
24658864	W	31.12.2018	0,000	397,000	397,000
Verbrauch (Einheiten)					2.778,000

Kostenaufstellung des gesamten Objektes

	Menge Fernwärme in kWh	Datum	Kosten in EUR	Zwischensumme in EUR	Gesamtsumme in EUR
Heizungsanlage					
Brennstoff					
Anlieferung Brennstoff	147.756,000	31.12.2018	16.513,82		
Verbrauch	147.756,000			16.513,82	
Weitere Heizungsbetriebskosten					
Betriebsstrom 2.00% von Brennstoff			330,28		
Verbrauchserfassung			892,74	1.223,02	
Zusatzkosten Heizung					
Wartung		06.11.2018	130,89		
Schornsteinfeger		31.12.2018	748,72		
Gerätemiete Heizung			408,77	1.288,38	
Gesamtkosten Heizungsanlage				19.025,22	19.025,22
Zu verteilende Gesamtkosten					19.025,22

BRECHNUNG

We measure for all sessions

1. “Conversion” (yes/no)

Variable	Group	Group size	Value
Conversion rate with ML support	Stage 1	101	0.48
	Stage 2	109	0.48
Conversion rate without ML support	Stage 1	116	0.47
	Stage 2	252	0.23
	Overall	368	0.30

Distinguish between groups:

Beta stage vs. final prototype stage

Manual vs. ML-supported data entry

n Can Guide AI Development for a Consumer-Oriented Self-service Portal.

ML-based Systems

‘The Loop’

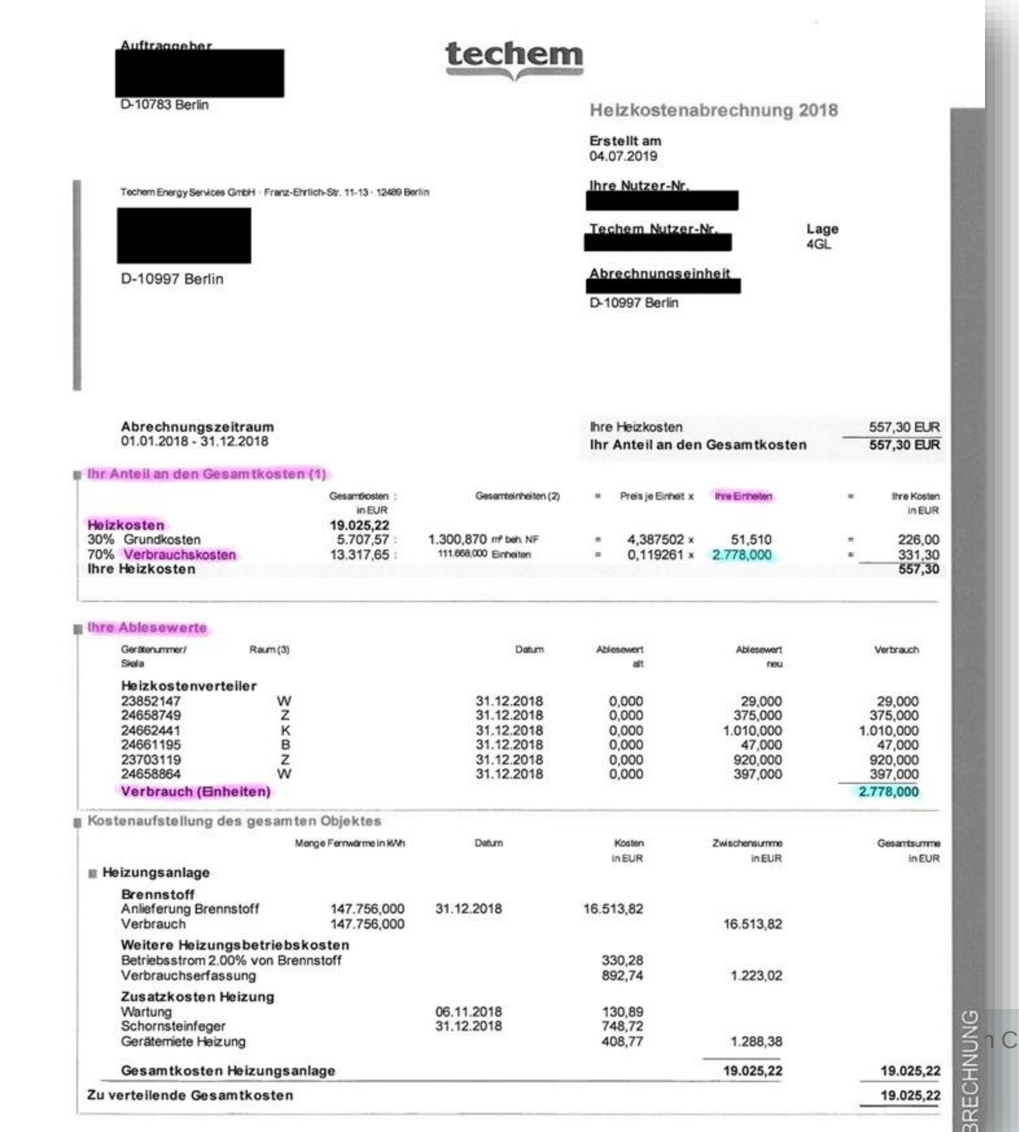
Roles within the

Use case / case

User research

Measure user su
and ML performa

Measuring user success and ML performance



Auftraggeber
[redacted]
D-10783 Berlin

techem

Heizkostenabrechnung 2018

Erstellt am
04.07.2019

Techem Energy Services GmbH · Franz-Ehrlich-Str. 11-13 · 12489 Berlin

[redacted]
D-10997 Berlin

Ihre Nutzer-Nr.
[redacted]

Techem Nutzer-Nr.
[redacted]

Lage
4GL

Abrechnungseinheit
[redacted]
D-10997 Berlin

Abrechnungszeitraum
01.01.2018 - 31.12.2018

Ihre Heizkosten	557,30 EUR
Ihr Anteil an den Gesamtkosten	**557,30 EUR**

Ihr Anteil an den Gesamtkosten (1)

	Gesamtkosten in EUR	:	Gesamteinheiten (2)	=	Preis je Einheit	x	Ihre Einheiten	=	Ihre Kosten in EUR
Heizkosten	**19.025,22**								
30% Grundkosten	5.707,57	:	1.300,870 m² beh. NF	=	4,387502	x	51,510	=	226,00
70% Verbrauchskosten	13.317,65	:	111.668,000 Einheiten	=	0,119261	x	2.778,000	=	331,30
Ihre Heizkosten									**557,30**

Ihre Ablesewerte

Gerätenummer/ Skala	Raum (3)	Datum	Ablesewert alt	Ablesewert neu	Verbrauch
Heizkostenverteiler					
23852147	W	31.12.2018	0,000	29,000	29,000
24658749	Z	31.12.2018	0,000	375,000	375,000
24662441	K	31.12.2018	0,000	1.010,000	1.010,000
24661195	B	31.12.2018	0,000	47,000	47,000
23703119	Z	31.12.2018	0,000	920,000	920,000
24658864	W	31.12.2018	0,000	397,000	397,000
Verbrauch (Einheiten)					**2.778,000**

Kostenaufstellung des gesamten Objektes

	Menge Fernwärme in kWh	Datum	Kosten in EUR	Zwischensumme in EUR	Gesamtsumme in EUR
Heizungsanlage					
Brennstoff					
Anlieferung Brennstoff	147.756,000	31.12.2018	16.513,82		
Verbrauch	147.756,000			16.513,82	
Weitere Heizungsbetriebskosten					
Betriebsstrom 2.00% von Brennstoff			330,28		
Verbrauchserfassung			892,74	1.223,02	
Zusatzkosten Heizung					
Wartung		06.11.2018	130,89		
Schornsteinfeger		31.12.2018	748,72		
Gerätemiete Heizung			408,77	1.288,38	
Gesamtkosten Heizungsanlage				**19.025,22**	**19.025,22**
Zu verteilende Gesamtkosten					**19.025,22**

BRECHNUNG

We measure for all sessions

1. “Conversion” (yes/no)

And for the ML-enabled mode for each bill:

2. Number of identified target regions
3. Number of “correctly” extracted target values, i.e. values that the users left unchanged during post correction

Distinguish between groups:

Beta stage vs. final prototype stage

Manual vs. ML-supported data entry

ML-based Systems

‘The Loop’

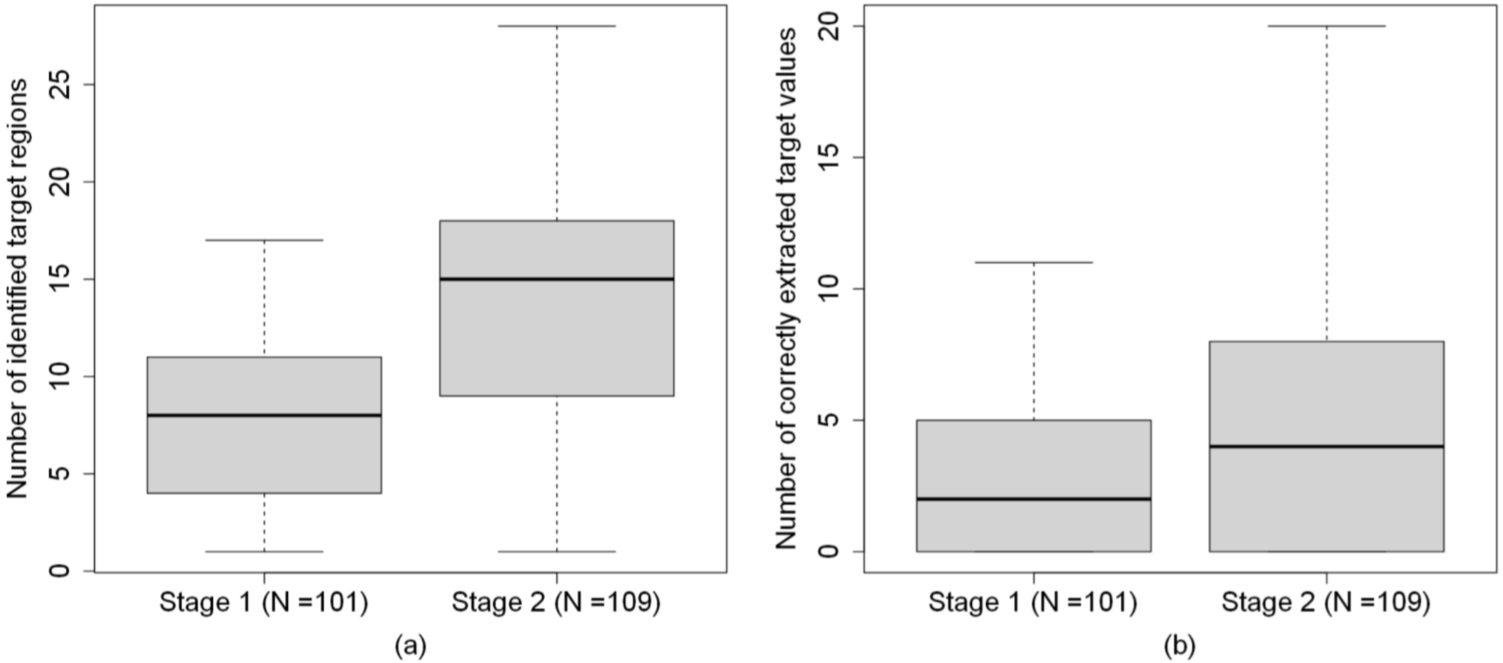
Roles within the loop

Use case / case study

User research

Measure user success and ML performance

Measuring user success and ML performance



Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

ML-based Systems

'The Loop'

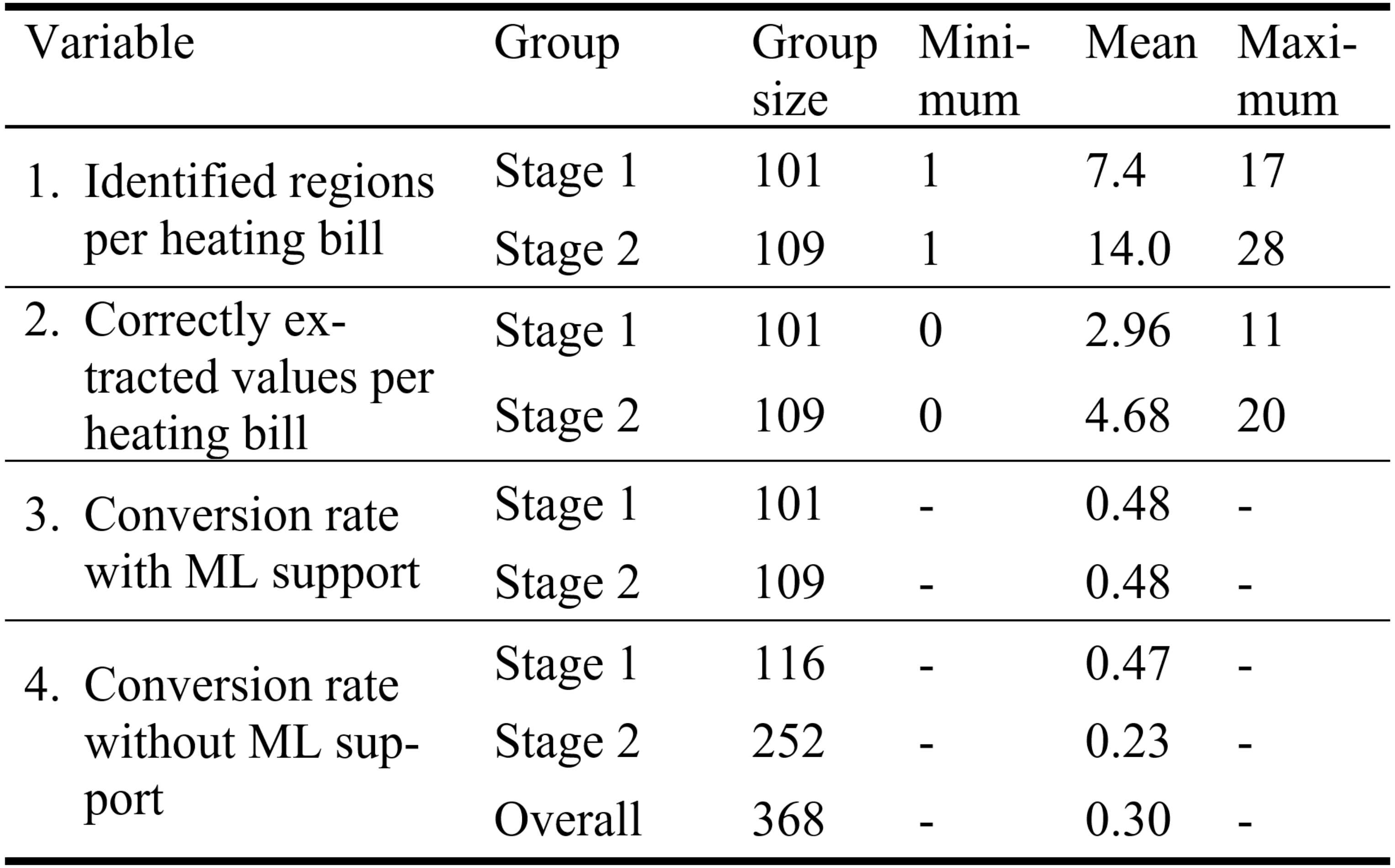
Roles within the loop

Use case / case study

User research

Measure user success and ML performance

Measuring user success and ML performance



Variable	Group	Group size	Minimum	Mean	Maximum
1. Identified regions per heating bill	Stage 1	101	1	7.4	17
	Stage 2	109	1	14.0	28
2. Correctly extracted values per heating bill	Stage 1	101	0	2.96	11
	Stage 2	109	0	4.68	20
3. Conversion rate with ML support	Stage 1	101	-	0.48	-
	Stage 2	109	-	0.48	-
4. Conversion rate without ML support	Stage 1	116	-	0.47	-
	Stage 2	252	-	0.23	-
	Overall	368	-	0.30	-

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

ML-based Systems

'The Loop'

Roles within the loop

Use case / case study

User research

Measure user success and ML performance

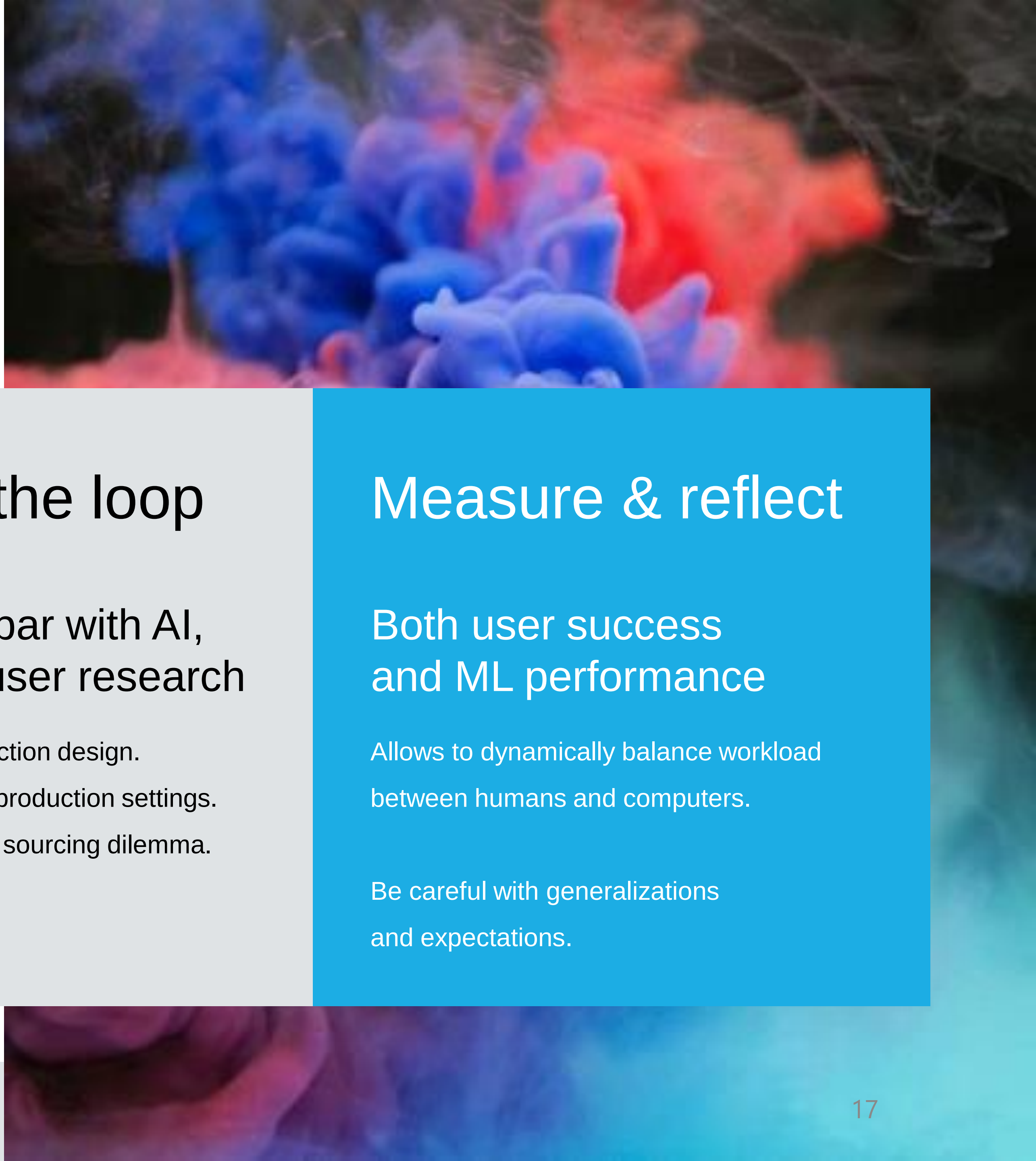
Measuring user success and ML performance

Limitations & challenges

1. Overall performance of our ML component / prototype
2. Small sample size, small data set (unstructured data)
3. Sample collected during ongoing user research, hence includes incentivized users
   - Users' post-correction of target values cannot be trusted
4. Hard to assess ML accuracy "on stage" with user-provided data
5. ML-component does not learn directly from user interactions

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal. *Culture and Computing. HCII 2022.*

Summary



Case study

Consumer-oriented self-service portal

To check **heating cost bills** for potential cost and energy savings.

Part of the services provided by:
https://www.co2online.de/

Users in the loop

Not only to spar with AI, but through user research

Leads to better interaction design.

Allows to evaluate in production settings.

Helps overcome data sourcing dilemma.

Measure & reflect

Both user success and ML performance

Allows to dynamically balance workload between humans and computers.

Be careful with generalizations and expectations.

Binder, F. et al. (2022). Putting Users in the Loop: How User Research Can Guide AI Development for a Consumer-Oriented Self-service Portal.
Culture and Computing. HCII 2022.

The end.

Light-weight typology of ML use cases

Aspect	Categories
Target group	**Consumers**, businesses, educational institutions, public administration, etc.
Type of data	**Unstructured data**, structured data
Size of data set	**Small data set**, large data set
Type of use case	Quality assurance in production; chatbots; error reduction, route optimization; process automation; predictive maintenance; **automated transaction processing** (damage notification or similar); **customer self-service;** supply chain optimization; intelligent / smart product development

Needs to be validated and refined.

Based on [25,27].