Artificial Intelligence

Algorithms and Applications with Python

Chapter 10

Dr. Dominik Jung
dominik.jung42@gmail.com

Outline

10	Building Productive AI-based Systems

10.1	Why AI Projects Fail
10.2	Human AI Interaction
10.3	Ethics and Trust in AI
10.4	Explainable and Understandable AI (XAI)
10.5	How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge

Image source: ↗ Pixabay (2019) / ↗ CC0

► **What you will learn:**
- Why many AI-based Information systems and intelligent agents fail in practice and how to tackle common problems in implementing AI-based information systems and intelligent agents
- Outlook and Future steps if you are interested in an AI-job

► **Duration:**
- 180 min

► **Relevant for Exam:**
- 10.1-10.3

- In the last chapters, we have discussed how to build AI their potential for automation. But is AI really better than humans in decision-making?

- Yes! There is overhelming evidence that artificial intelligence, clearly outperforms human decision-making:
  - **Paul Meehl**: „Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evicence"
  - **Daniel Kahneman:** „Thinking Fast and Slow", or „Noise"
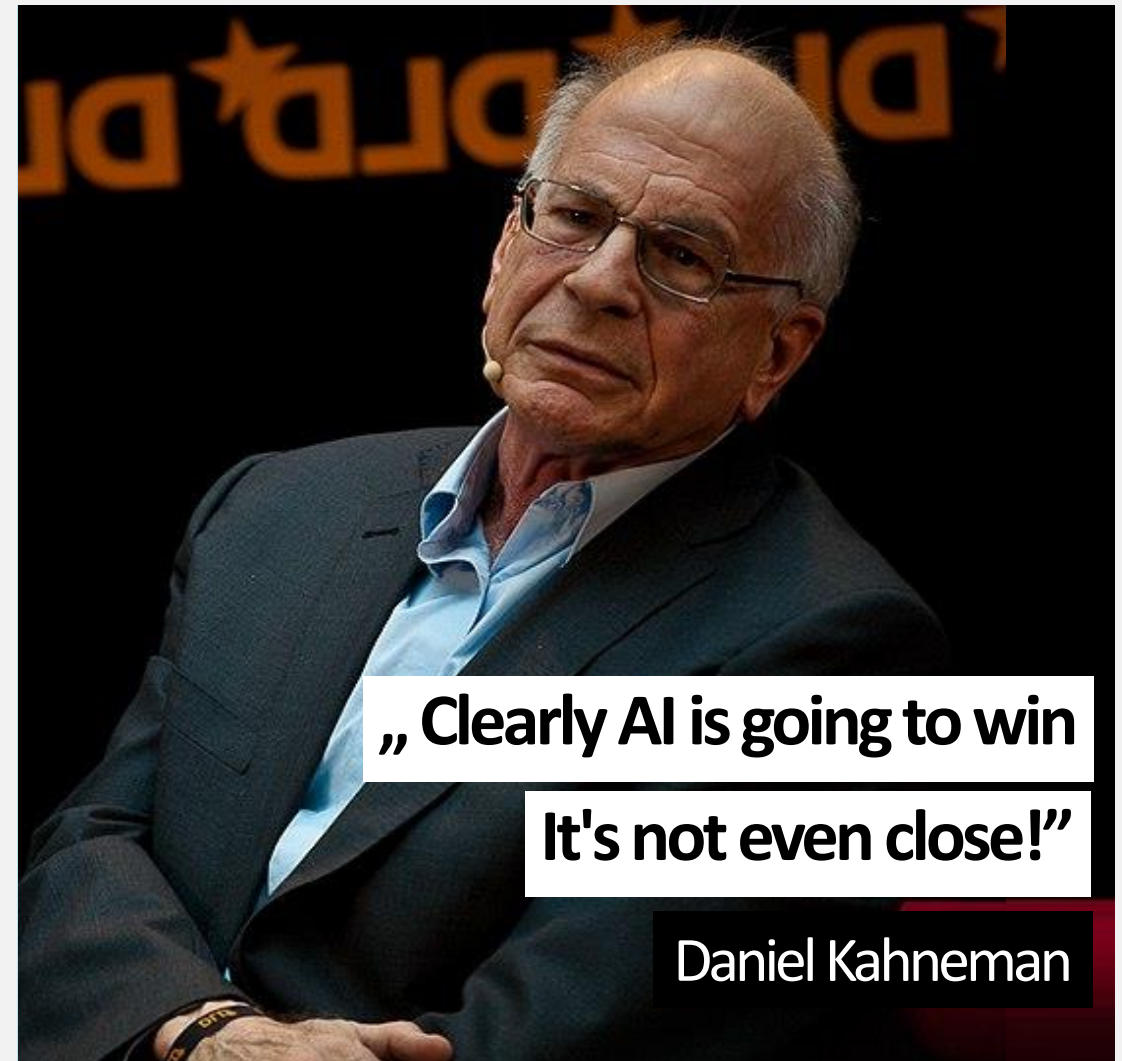
- But does it really succeed?

„ **Clearly AI is going to win**

It's not even close!"

Daniel Kahneman

Image source: ↗Daniel Kahneman (2009) by nrkbeta from Wikimedia / ↗CC BY-SA 3.0

IEEE SPECTRUM

Feature | Biomedical | Diagnostics

02 Apr 2019 | 15:00 GMT

How IBM Watson Overpromised and Underdelivered on AI Health Care

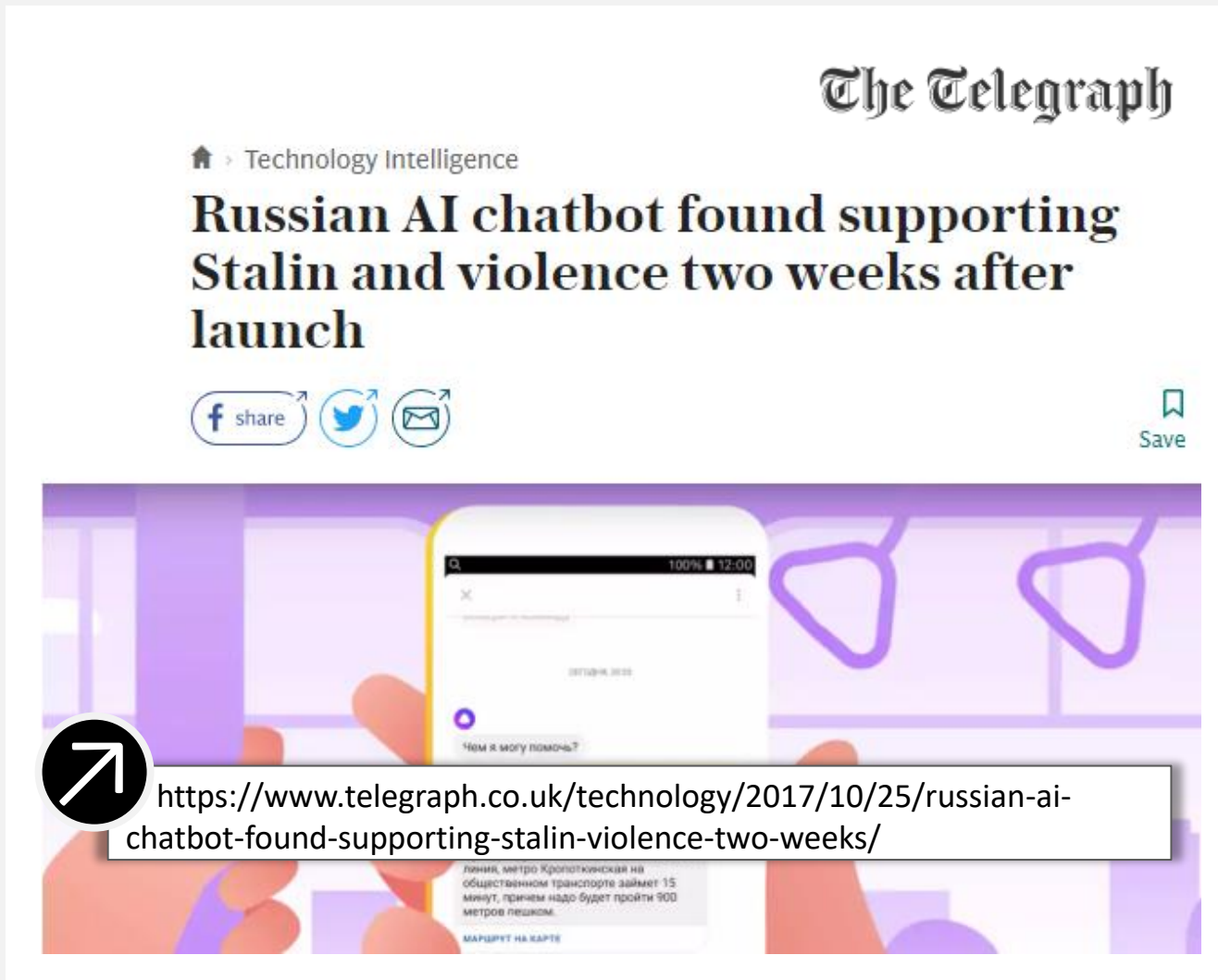After its triumph on *Jeopardy!*, IBM's AI seemed poised to revolutionize medicine. Doctors are still waiting

By **Eliza Strickland**

https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

- 62 Million dollars has been spend by University of Texas before the contract was canceled

- Chamath Palihapitiya: "Watson is a joke, just to be honest. I think that IBM is excellent at **using their sales and marketing infrastructure to convince people** who have **asymmetrically less knowledge to pay** for something."

The Telegraph

🏠 › Technology Intelligence

Russian AI chatbot found supporting Stalin and violence two weeks after launch

https://www.telegraph.co.uk/technology/2017/10/25/russian-ai-chatbot-found-supporting-stalin-violence-two-weeks/

- Users of the "Alice" assistant have reported it responding positively to questions about domestic violence and saying that "enemies of the people" must be shot

- Supports wife-beating, child abuse and suicide

GOOGLE MAPS IN BERLIN

Wenn aus Aussetzfahrten eine neue U-Bahn-Linie wird

Kartendienste sind für Touristen wie auch Ortskundige längst eine willkommene Hilfe. Doch manchmal gibt es größere Fehler. In Berlin werden beispielsweise einige Kleinprofil-Linien falsch gerendert. Dabei werden betriebliche Besonderheiten dargestellt.

Von Andreas Sebayang

9. Juli 2019, 11:12 Uhr

https://www.golem.de/news/google-maps-in-berlin-wenn-aus-aussetzfahrten-eine-neue-u-bahn-linie-wird-1907-142415.html

- Google determines the routes of tram lines through AI.

- However, this leads sometimes to obviously wrong results: data from public transport was misinterpreted and Google Maps had for some days new underground lines in Berlin.

"60% of models developed to operationalize them were never actually operationalized"

Gartner Data Science Team Survey of January 2018

ginablaber
@ginablaber

Follow

The story of enterprise Machine Learning: "It took me 3 weeks to develop the model. It's been >11 months, and it's still not deployed."
@DineshNirmalIBM #StrataData #strataconf

10:19 AM - 7 Mar 2018

Why 85% of AI projects fail

TechRepublic.

The Myth of Agile AI/Machine Learning in the Enterprise

Eric Broda   Follow
Jul 31, 2019 · 13 min read ★

M

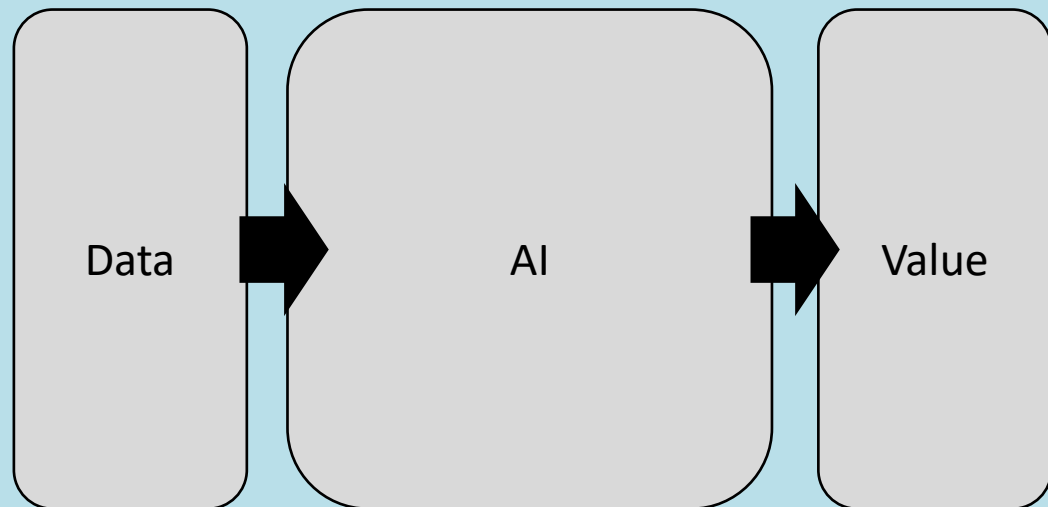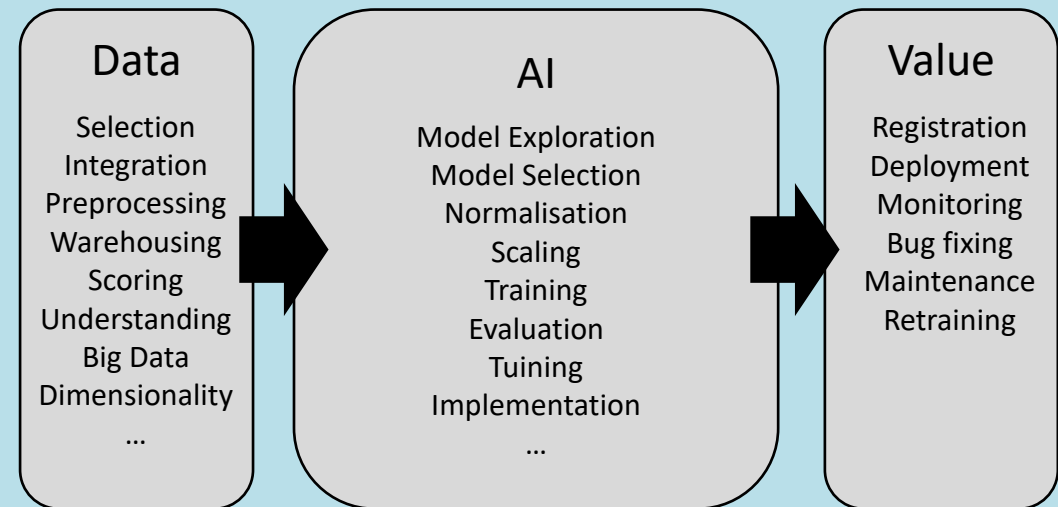Are AI applications failing to live up to hype in enterprise adoption?

ComputerWeekly.com

?

Why do some AI projects fail in real life?

10.1 Problem: Building Productive and Usefull AI Systems is Hard

What companies think AI looks like

Data → AI → Value

What it actually is

Data
Selection
Integration
Preprocessing
Warehousing
Scoring
Understanding
Big Data
Dimensionality
…

AI
Model Exploration
Model Selection
Normalisation
Scaling
Training
Evaluation
Tuining
Implementation
…

Value
Registration
Deployment
Monitoring
Bug fixing
Maintenance
Retraining

! Legal issues

Ethics

Security

1 **Definition of the Problem**
- What should be automated?
- Which kind of AI problem do I have?
- What options are available?

2 **Use Case Environment**
- Who is my customer? What are the users' characteristics of my AI-based Information System?
- What is the goal of the Use Case?

3 **IT Environment**
- What are the prerequisites of the chosen tool?
- What type of IT-Infrastructure is available?

The first step is to find out which class of AI-based Information System is required, and which kind of AI problem has to be solved:

Some example goals of use cases might be…

- Visualizing key characteristics of a dataset to understand the drivers of a specific problem (e.g. specific dealers have higher costs then other dealers)

- Create a proof of concept of decision-support systems like a dashboard to make management-decisions (e.g. make sales forcasts on specific data)

Specify the objective

The objective of the AI-project has not been specified. This leads to problems on a general level.

1. What specific task should the model automate?

2. Which data should be used/can be used for the task? How reliable is the data?

- Understand the problem from the perspective of the user (subject yourself to the problem)

- Integrate the User in the development process

- Do not underestimate the need for accurate data

Understand the problem

AI-based information systems have low usability

1. How does the user interact with the AI-based information system?

2. How can it actually help the user to be more efficient?

- Prototype and iterate on the user experience using mockups or sketches

- Understand the -real- user behavior: user studies, workshops, wizard of oz experiments etc.

- Rely on established guidelines

Design (user-)specific

AI-based information systems have low coherence and often are too general instead of user specific tools

1. What information should we expose to the user?

2. How should we expose information to the user?

- Develop a shared language with your project stakeholders

- Start with simple models, deploy a productive baseline model as soon as possible

- Deliver value incrementally and quickly

Adapted from Bernardi, L et al. (2019), Google (2019)

- Start with informational one-on-one interviews with the end users to get a feeling of their perspective

- In this phase do not start suggesting or developing (!) a solution, you are simply trying to understand the problem

- Nevertheless, it can make sense to communicate the capabilities of AI-based information systems

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

10.1 Subject Yourself to the Problem

- After the informational interviews, perform the task that should be automated manually (or the problem that should be solved)

- Do not jump right into modelling!

- If you do the task, pay attention to how you and the key users solve the task. This gives you a first impression of first features, that might be important.

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

- Now, you can start sketching out first possible solutions of the AI-based information system you plan to develop

- Iterate on the user experience using prototypes, sketches and mockups to communicate first possible solutions

- Ask questions and find the motivations behind solutions (!)

- To test the AI module use wizard-of-oz experiments

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

- To test your mockups with 1-2 selected users will NOT help you

- Even during this early feedback stages, look for a wide range of different test users with different characteristics:

? What are typical types of characteristics you should consider?

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

1

Definition of the Problem
- What should be automated?
- Which kind of AI problem do I have?
- What options are available?

2

Use Case Environment
- Who is my customer? What are the users' characteristics of my AI-based Information System?
- What is the goal of the Use Case?

3

IT Environment
- What are the prerequisites of the chosen tool?
- What type of IT-Infrastructure is available?

Having a clearly-defined goal usually brings with it other prerequisites on a tool, e.g.:

- How does the **operative environment** in which the final outcome of the project is to be deployed look like?
- Is it a "Big Data Use Case" (volume and velocity of data flow)?
- Price (open source?) Availability of APIs to other tools
- Ease of use (e.g. presence of a GUI)
- Computational efficiency
- Availability of technical support/online documentation/community

- Do not fear to over-communicate: Talk, talk, talk

- Discuss and present your ideas and progress often in short daily/weekly/etc. to ensure that you are still going in the right direction

- Discuss model metrics / performance measures to illustrate the progress of the model. This allows your project partner to present and justify the project in the organization

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

Google Developer Guides:
↗https://developers.google.com

Rules of Machine Learning:

Best Practices for ML Engineering

Martin Zinkevich

- **Deploy simple models in the spirit of winning by shipping**

- **Make quick iterations through the outermost development loop**

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

- Deploy a baseline model on production data as soon as possible

- Deploying your model on production data can be enlightening

- As a countermeasure, it's often a good idea to deploy a simple model on production data as soon as possible

- Deliver value incrementally and quickly

- Measure time to results, not results. Sometimes it can be tricky to get into the mindset of delivering value quickly.

Based on Bernardi, L et al. (2019); Google (2019); Zinkevich, M (2019)

1 **Definition of the Problem**
- What should be automated?
- Which kind of AI problem do I have?
- What options are available?

2 **Use Case Environment**
- Who is my customer? What are the users' characteristics of my AI-based Information System?
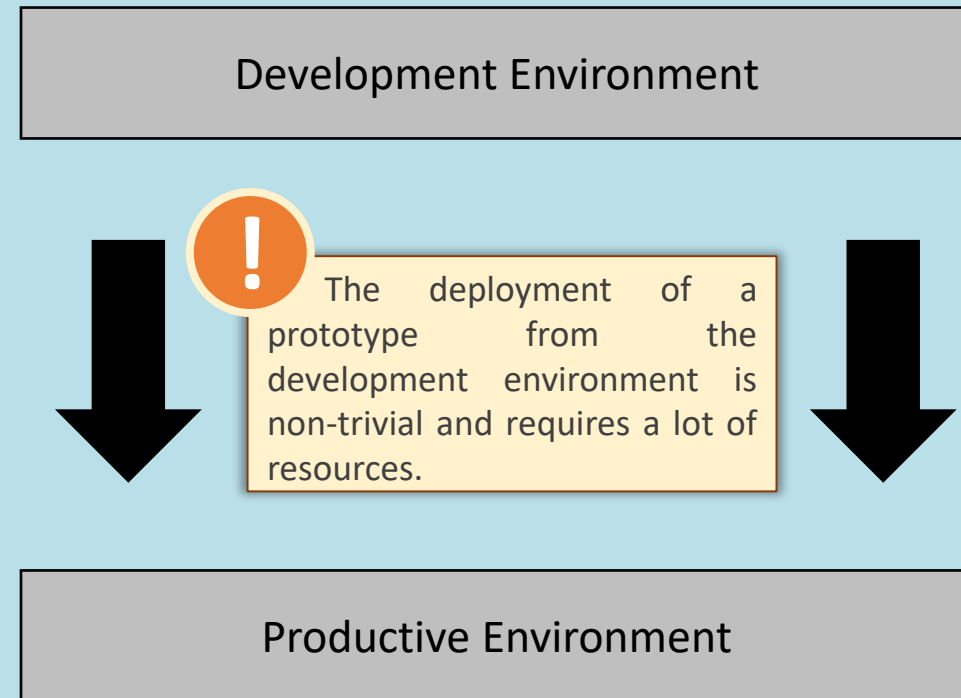- What is the goal of the Use Case?

3 **IT Environment**
- What are the prerequisites of the chosen tool?
- What type of IT-Infrastructure is available?
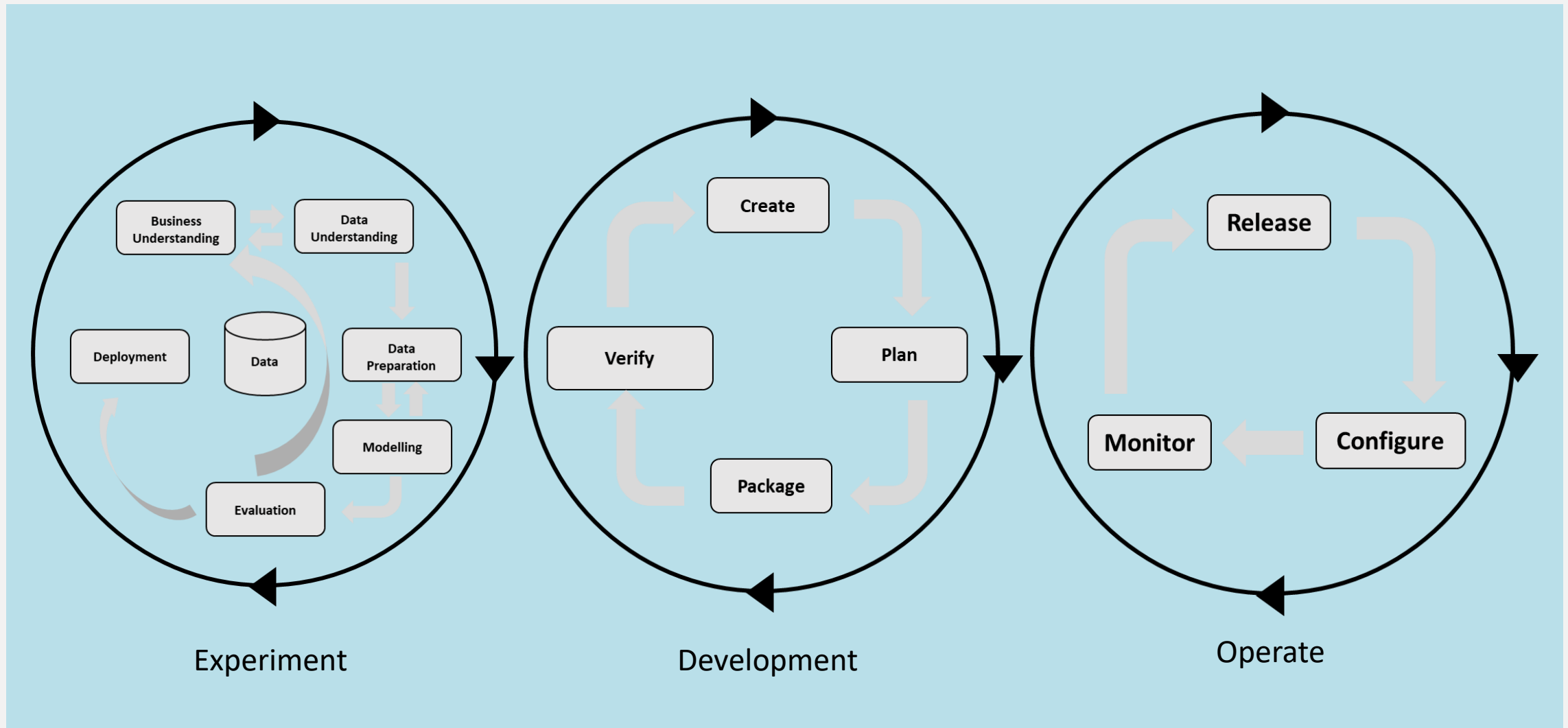
The first step is to find out what the Toolbox of the respective department is:

- What tools are even available?

- How fast can additional licenses for proprietary software be obtained and how many are required?

- Is there the option of introducing new tools or will the approval process take too long?
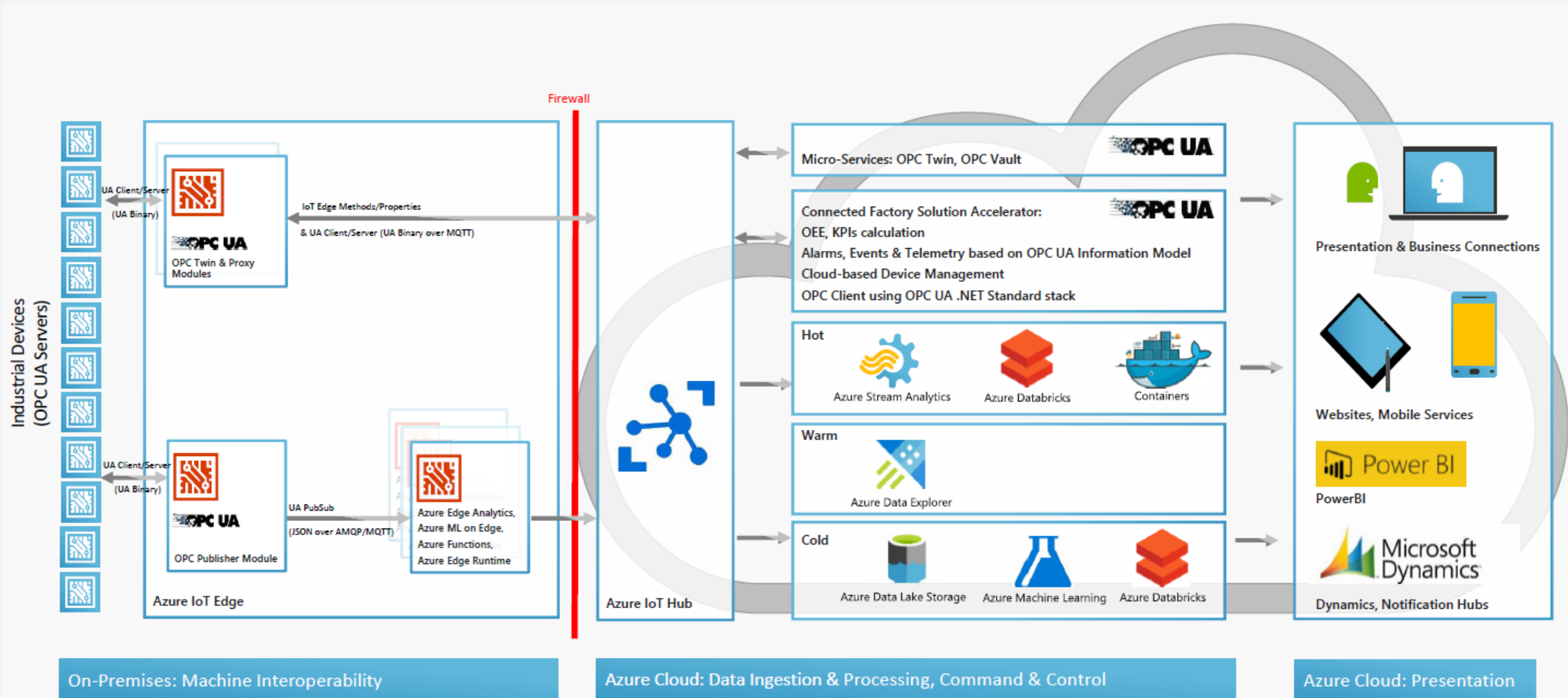
- Most projects consist of several phases and aspects, which together make up an AI-based IS's life cycle.

- The life cycle spans the very inception of a data science use case up until its productive rollout

Development Environment

The deployment of a prototype from the development environment is non-trivial and requires a lot of resources.

Productive Environment

Experiment

Development

Operate

ML + App Dev Process

Business Environ-mental Factors

- Digitalization
- Globalization
- Consumer demands
- Government regulations
- Markets & competition
- …

Pressures →

Opportunities →

Organization Responses

- Strategy
- Partners' collaboration
- Real-time response
- Agility
- Increased productivity
- New vendors
- New business models
- …

→

Decisions and Support

- Analysis
- Predictions
- Decisions

Integrated computerized decision support

Business intelligence / analytics

Information Systems
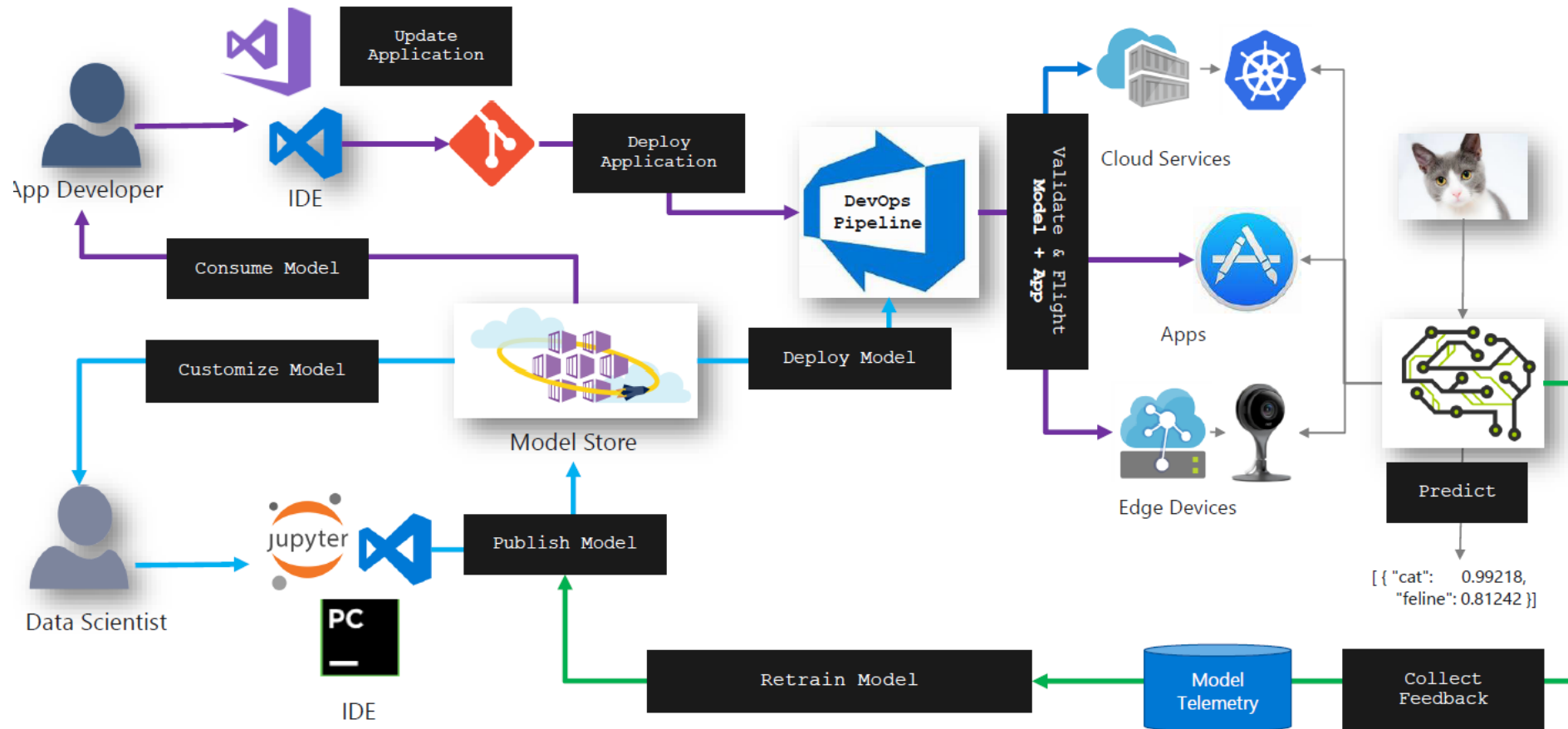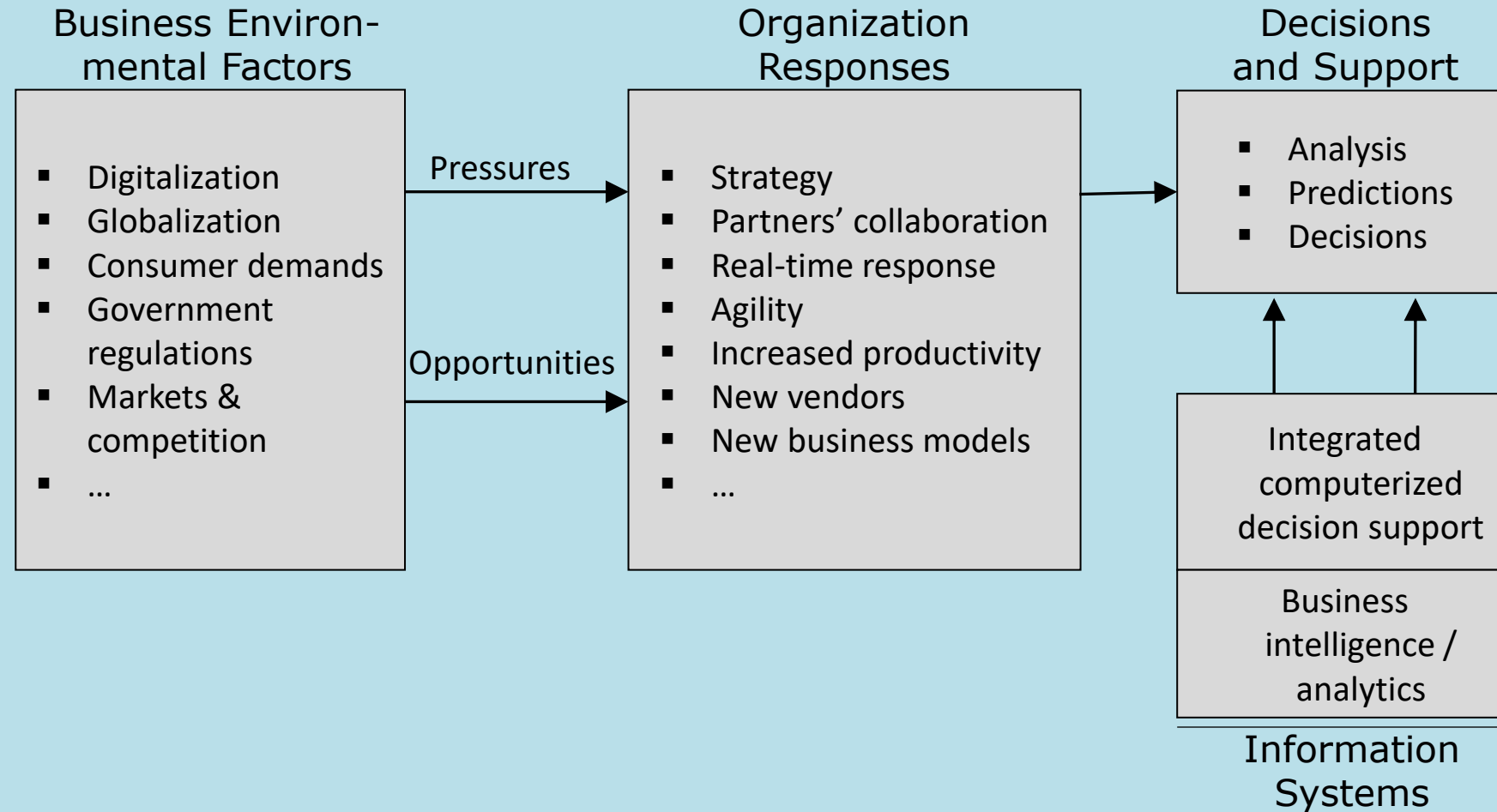
Based on Based on Sharda et al. (2014)
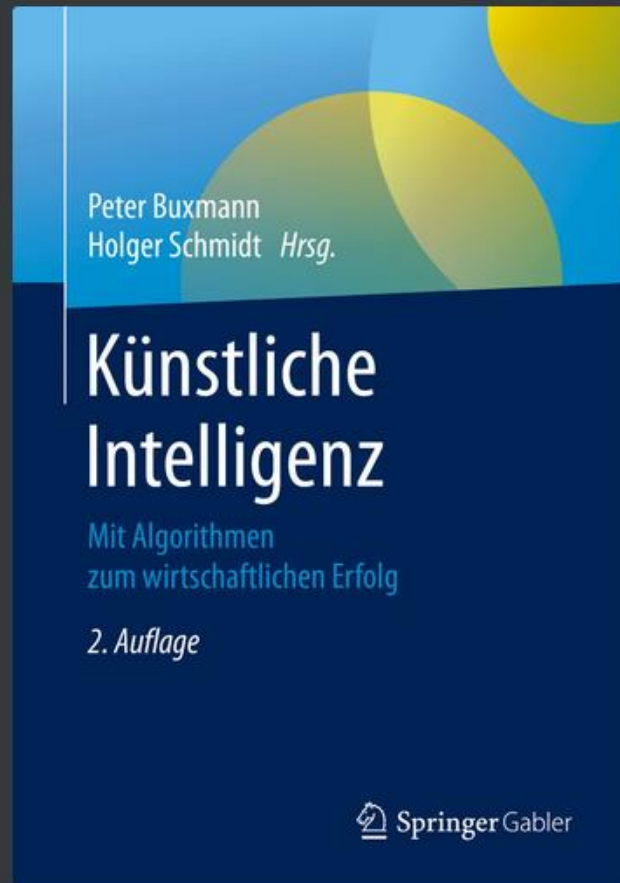
Künstliche Intelligenz wird als Basistechnologie ähnlich wie Elektrizität viele Unternehmen und Arbeitsplätze grundlegend verändern. Nach einer allgemein verständlichen Einführung in die technischen Grundlagen und ökonomischen Effekte zeigen zehn Unternehmen in der 2. Auflage unseres Buches, wie sie Algorithmen einsetzen, um Routinejobs zu automatisieren, aus Daten Erkenntnisse abzuleiten und neue Geschäftsmodelle zu entwickeln. Mit dem Buch wollen wir KI entmystifizieren, für Unternehmen handhabbar und für Interessierte verständlich machen. Wir wollen Entscheidern:innen Mut machen, sich mit dieser essentiellen Technologie stärker zu beschäftigen, damit der Vorsprung der amerikanischen und chinesischen Wettbewerber nicht noch größer wird. Mit Beiträgen von:

- Commerzbank: Künstliche Intelligenz in Banken
- Volkswagen: KI-Innovationen über das autonome Fahren hinaus
- Otto: Data Science im Online-Handel
- SAP: Effiziente Prozesse mit Künstlicher Intelligenz
- Microsoft: Künstliche Intelligenz in der Cloud
- Serviceware: Künstliche Intelligenz im Enterprise Service Management
- Software AG: Intelligentes IOT: Erkenntnisse aus IOT-D... ch Machine Learning
- IBM: Intelligente KI-Lösungen
- Empolis: Mit KI immer die richtigen Entscheidungen treffen
- Heraeus: KI-gestützte visuelle Inspektion in der Elektronik...

Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg (↗ki-business.de)

☑ Previous Exam Task!

Your turn!

Task

What are the three main problems an AI-designer should have in mind when he/she designs AI-based information systems? Choose an example and explain it.

Outline

10	Building Productive AI-based Systems

10.1	Why AI Projects Fail
10.2	Human AI Interaction
10.3	Ethics and Trust in AI
10.4	Explainable and Understandable AI (XAI)
10.5	How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge

Image source: ↗ Pixabay (2019) / ↗ CC0

►**What you will learn:**
- Why many AI-based Information systems and intelligent agents fail in practice and how to taggle common problems in implementing AI-based information systems and intelligent agents
- Outlook and Future steps if you are interested in an AI-job

►**Duration:**
- 180 min

►**Relevant for Exam:**
- 10.1-10.3

10.2 Guidelines for Human-AI Interaction from Microsoft

- The Researcher propose 18 general applicable design guidelines for human-AI interaction.

- Guidelines are validated through multiple rounds of evaluation including a user study where this guideline is tested against other popular AI-infused products.
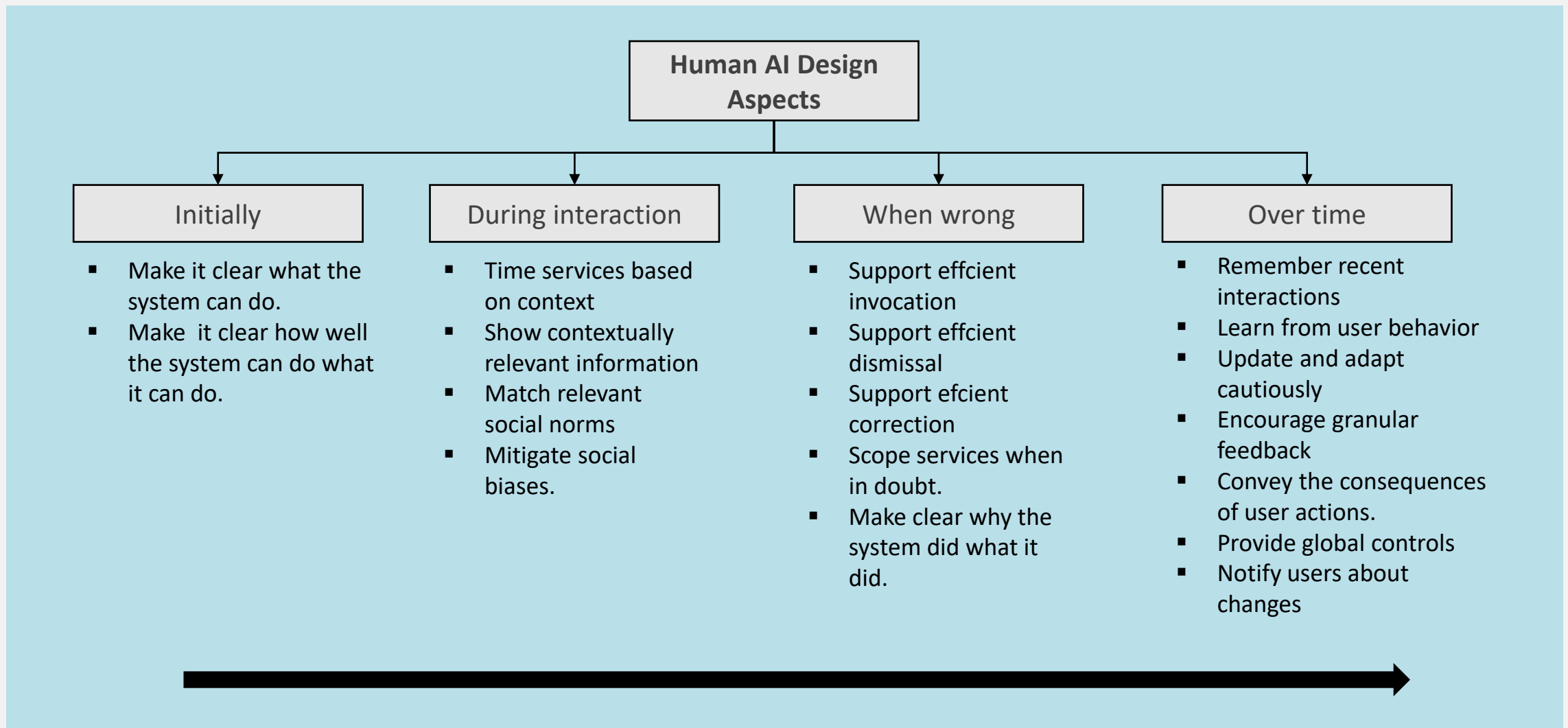
The following part of the lecture is mainly based on Microsoft's Best Practices for AI-Design ↗www.microsoft.com.

Adapted from Amershi, S et al. (2019)

		AI Design Guidelines	Example Applications of Guidelines
Initially	G1	**Make clear what the system can do.** Help the user understand what the AI system is capable of doing.	[Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day."
	G2	**Make clear how well the system can do what it can do.** Help the user understand how often the AI system may make mistakes.	[Music Recommenders, Product #1] "A little bit of hedging language: 'we think you'll like'."
During interaction	G3	**Time services based on context.** Time when to act or interrupt based on the user's current task and environment.	[Navigation, Product #1] "In my experience using the app, it seems to provide timely route guidance. Because the map updates regularly with your actual location, the guidance is timely."
	G4	**Show contextually relevant information.** Display information relevant to the user's current task and environment.	[Web Search, Product #2] "Searching a movie title returns show times in near my location for today's date"
	G5	**Match relevant social norms.** Ensure the experience is delivered in a way that users would expect, given their social and cultural context.	[Voice Assistants, Product #1] "[The assistant] uses a semi-formal voice to talk to you - spells out "okay" and asks further questions."
	G6	**Mitigate social biases.** Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.	[Autocomplete, Product #2] "The autocomplete feature clearly suggests both genders [him, her] without any bias while suggesting the text to complete."
When wrong	G7	**Support efficient invocation.** Make it easy to invoke or request the AI system's services when needed.	[Voice Assistants, Product #1] "I can say [wake command] to initiate."
	G8	**Support efficient dismissal.** Make it easy to dismiss or ignore undesired AI system services.	[E-commerce, Product #2] "Feature is unobtrusive, below the fold, and easy to scroll past...Easy to ignore."
	G9	**Support efficient correction.** Make it easy to edit, refine, or recover when the AI system is wrong.	[Voice Assistants, Product #2] "Once my request for a reminder was processed I saw the ability to edit my reminder in the UI that was displayed. Small text underneath stated 'Tap to Edit' with a chevron indicating something would happen if I selected this text."
	G10	**Scope services when in doubt.** Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.	[Autocomplete, Product #1] "It usually provides 3–4 suggestions instead of directly auto completing it for you"
	G11	**Make clear why the system did what it did.** Enable the user to access an explanation of why the AI system behaved as it did.	[Navigation, Product #2] "The route chosen by the app was made based on the Fastest Route, which is shown in the subtext."
Over time	G12	**Remember recent interactions.** Maintain short term memory and allow the user to make efficient references to that memory.	[Web Search, Product #1] "[The search engine] remembers the context of certain queries, with certain phrasing, so that it can continue the thread of the search (e.g., 'who is he married to' after a search that surfaces Benjamin Bratt)"
	G13	**Learn from user behavior.** Personalize the user's experience by learning from their actions over time.	[Music Recommenders, Product #2] "I think this is applied because every action to add a song to the list triggers new recommendations."
	G14	**Update and adapt cautiously.** Limit disruptive changes when updating and adapting the AI system's behaviors.	[Music Recommenders, Product #2] "Once we select a song they update the immediate song list below but keeps the above one constant."
	G15	**Encourage granular feedback.** Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.	[Email, Product #1] "The user can directly mark something as important, when the AI hadn't marked it as that previously."
	G16	**Convey the consequences of user actions.** Immediately update or convey how user actions will impact future behaviors of the AI system.	[Social Networks, Product #2] "[The product] communicates that hiding an Ad will adjust the relevance of future ads."
	G17	**Provide global controls.** Allow the user to globally customize what the AI system monitors and how it behaves.	[Photo Organizers, Product #1] "[The product] allows users to turn on your location history so the AI can group photos by where you have been."
	G18	**Notify users about changes.** Inform the user when the AI system adds or updates its capabilities.	[Navigation, Product #2] "[The product] does provide small in-app teaching callouts for important new features. New features that require my explicit attention are pop-ups."

Table 1: Our 18 human-AI interaction design guidelines, roughly categorized by when they likely are to be applied during interaction with users, along with illustrative applications (rated as "clearly applied" by participants) across products tested by participants in our user study.

Human AI Design Aspects

Initially	During interaction	When wrong	Over time
▪ Make it clear what the system can do. ▪ Make it clear how well the system can do what it can do.	▪ Time services based on context ▪ Show contextually relevant information ▪ Match relevant social norms ▪ Mitigate social biases.	▪ Support effcient invocation ▪ Support effcient dismissal ▪ Support efcient correction ▪ Scope services when in doubt. ▪ Make clear why the system did what it did.	▪ Remember recent interactions ▪ Learn from user behavior ▪ Update and adapt cautiously ▪ Encourage granular feedback ▪ Convey the consequences of user actions. ▪ Provide global controls ▪ Notify users about changes

Adapted from Amershi, S et al. (2019)

1

G1 - Make clear what the system can do.

- Help the user understand what the AI system is capable of doing
- **Example:** [Activity Trackers, Product #1] "Displays all the metrics that it tracks and explains how. Metrics include movement metrics such as steps, distance traveled, length of time exercised, and all-day calorie burn, for a day."
- **Example:** [Photo Organizers, Product #1] "We know the AI is able to detect and associate an image with a category, but the user does not know all the categories available."

Example: User Onboarding

Here you see <functionality>!

Try out to…

This button allows you to <function>

Adapted from Amershi, S et al. (2019)

G2 - Make clear how well the system can do what it can do.

2

- Help the user understand how often the AI system may make mistakes.
- **Example:** [Music Recommenders, Product #1] "A little bit of hedging language: 'we think you'll like'."
- **Example:** [Voice Assistants, Product #1] "Aside from the 'Hi, how can I help?', [the product] does not promise anything more. No expectation of quality is set."

Example: Message / show how often you used the functionality of the system

9 of 10
Warnings
correct

Number of correct cases

Adapted from Amershi, S et al. (2019)

G3 - Time services based on context.

3

- Time when to act or interrupt based on the user's current task and environment
- **Example:** [Autocomplete, Product #1] "Suggestions are always present when you might need them (whenever the keyboard is up)"
- **Example:** [Social Networks, Product #1] "If the user has not accessed [the product] in a while, the application will let the user know that there is something new to be explored - a story, video, etc."

Example: User-sensitive systems

KÜNSTLER*INNEN-RADIO BASIEREND AUF
Die drei ???

kle Wächter - Teil 38
Die drei ???

0:04   1:46

TURN OFF CAR VIEW

You can turn off car view until your next drive. Go to Settings to turn off car view permanently.
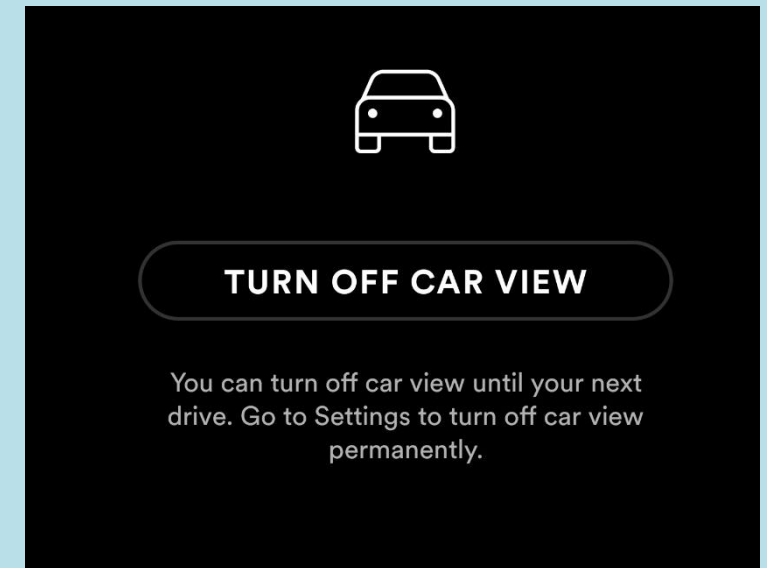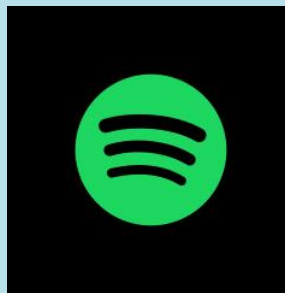
Adapted from Amershi, S et al. (2019)

10.2 AI Design Guideline: During interaction

G4 - Show contextually relevant information.

4

- Display information relevant to the user's current task and environment.
- **Example:** [Web Search, Product #2] "Searching a movie title returns show times near my location for today's date"
- **Example:** [Navigation, Product #1] "When I use [the product] for driving directions, it remembers where I parked my car. Next time when I open the app, it suggests routing me back to my car."

Example: Search a recent movie in the google search of your android phone and it returns playtimes of your local cinema.

HEUTE, 13. APR.

Nightlife >

21:00

The Grudge >
Horror · 1 Std., 34 Min.

23:45

Weitere Filme

Adapted from Amershi, S et al. (2019)

G5 - Match relevant social norms

5

- Ensure the experience is delivered in a way that users would expect, given their social and cultural context.
- **Example:** [Voice Assistants, Product #1] "[The assistant] uses a semiformal voice to talk to you - spells out "okay" and asks further questions."
- **Example:** [Navigation, Product #1] "If you select walking, the AI avoid(s) busy roads and searches for trails."

Example: Example Mastercard chatbot interface

(I know like most chatbots this is no 100% AI example)

YES, PLEASE

NOPE, UNDO

YES, PLEASE

You've send Trucy 5$.

Thanks. See you

Ok! Until soon.

Message MasterCard

Adapted from Amershi, S et al. (2019)

6

G6 - Mitigate social biases.

- Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.
- **Example:** [Web Search, Product #2] "a search for CEO or Doctor shows somewhat diverse people in the resulting images...The images are pretty diverse in terms of gender and ethnicity, although still lack in some respects such as disability"
- **Example:** [Autocomplete, Product #2] "The autocomplete feature clearly suggests both genders [him, her] without any bias while suggesting the text to complete."

Example: Yandex Assistant
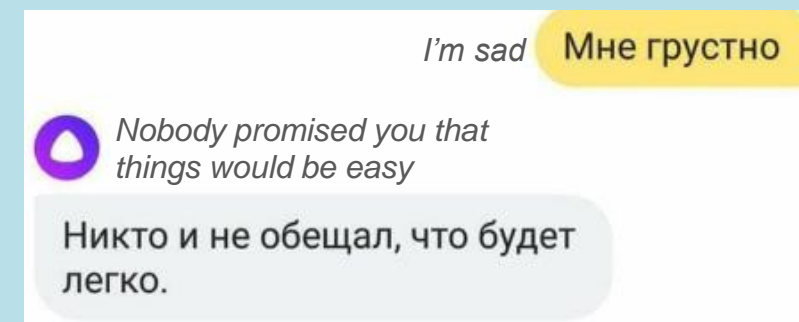
I'm sad Мне грустно

Nobody promised you that things would be easy

Никто и не обещал, что будет легко.

Adapted from Amershi, S et al. (2019)

G7 - Support efficient invocation.

7

- Make it easy to invoke or request the AI system's services when needed.
- **Example:** [Voice Assistants, Product #1] "I can say [wake command] to initiate."
- **Example:** [E-commerce, Product #1] "In addition to the system giving you recommendations as you browse, you can go to your "Browsing history > Manage history > More like this" to get recommendations specific to a particular product."

Example: Google Work

If you add your workplace and your home, Google warns you if there are any unexpected events (e.g. traffic jam)

Furthermore, it adds your standard ways to the app interface
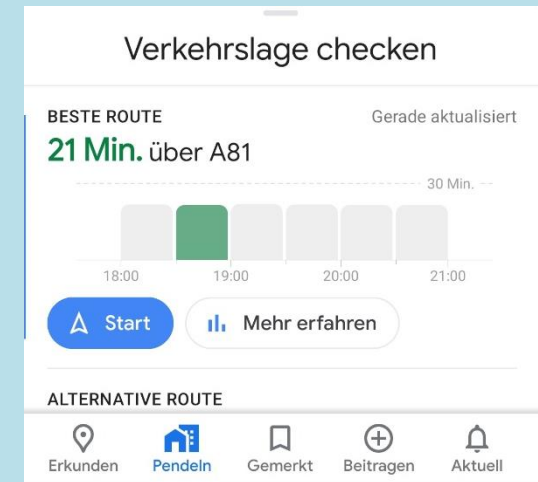
Adapted from Amershi, S et al. (2019)

G8 - Support efficient dismissal

8

- Make it easy to dismiss or ignore undesired AI system services.
- **Example:** [E-commerce, Product #2] "Feature is unobtrusive, below the fold, and easy to scroll past...Easy to ignore."
- **Example:** [Voice Assistants, Product #1] "I can say "nevermind" to dismiss it once I have said [wake command]. I can also just not say anything and it stops listening."

Stop or pause:
- "Alexa, stop" or,
- "Alexa, shut up."

Mute or unmute:
- "Alexa, mute" or,
- "Alexa, unmute."

Adapted from Amershi, S et al. (2019)

G9 - Support efficient correction.

9

- Make it easy to edit, refine, or recover when the AI system is wrong.
- **Example:** [Navigation, Product #1] "If [the product] is wrong about where I parked my car, it provides an easy way to edit the location by dragging on the map."
- **Example:** [Web Search, Product #2] "automatically 'corrects' spelling errors, etc. but gives option at top to return to query as originally typed...Notes that the query had been corrected and is one click to revert back to original"
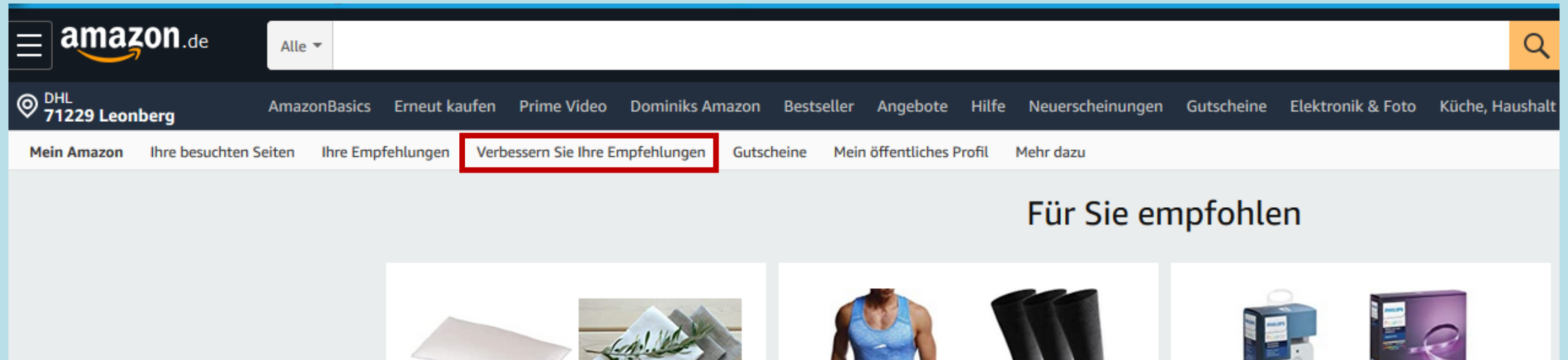
Adapted from Amershi, S et al. (2019)

G10 - Scope services when in doubt.

10

- Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.
- **Example:** [Navigation, Product #1] "If more than one line takes the same route the user can choose between the preferred line."
- **Example:** [Autocomplete, Product #1] "It usually provides 3-4 suggestions instead of directly auto completing it for you"

I would like a romantic place for Italian food near my office

→
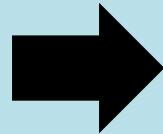
Hey, I found these Italian restaurants which reviews say are romantic close to your work

Adapted from Amershi, S et al. (2019)

G11 - Make clear why the system did what it did.

11

- Enable the user to access an explanation of why the AI system behaved as it did.
- **Example:** [E-commerce, Product #1] "Clicking "Why recommended" explains why they have recommended that particular item to you."
- **Example:** [Music Recommenders, Product #2] "I think this applies because each of recommendation has some information as to which songs are displayed on it - similar to the song, from the same artist, from the same album etc."

Adapted from Amershi, S et al. (2019)

12

G12 - Remember recent interactions.

- Maintain short term memory and allow the user to make efficient references to that memory.
- **Example:** [Navigation, Product #1] "Opening the app shows a list of recent destinations, as well as allows you to access "favorite" locations."
- **Example:** [Voice Assistants, Product #1] "[The assistant] seems to remember conversation context at least one command back. When asked "[wake command], what's the reminder?" she announces the last unheard remind me."

Resume the last played audiobook: "Alexa, resume my book."

audible
Ein amazon Unternehmen

Adapted from Amershi, S et al. (2019)

G13 - Learn from user behavior.

13

- Personalize the user's experience by learning from their actions over time.
- **Example:** [Music Recommenders, Product #2] "I think this is applied because every action to add a song to the list triggers new recommendations."
- **Example:** [Email, Product #1] "(My guess is) the system learns from what previous emails have attracted more attention from me (i.e. longer/more frequent reply, reading time taken, longer email threads, etc.) and infer email importance."
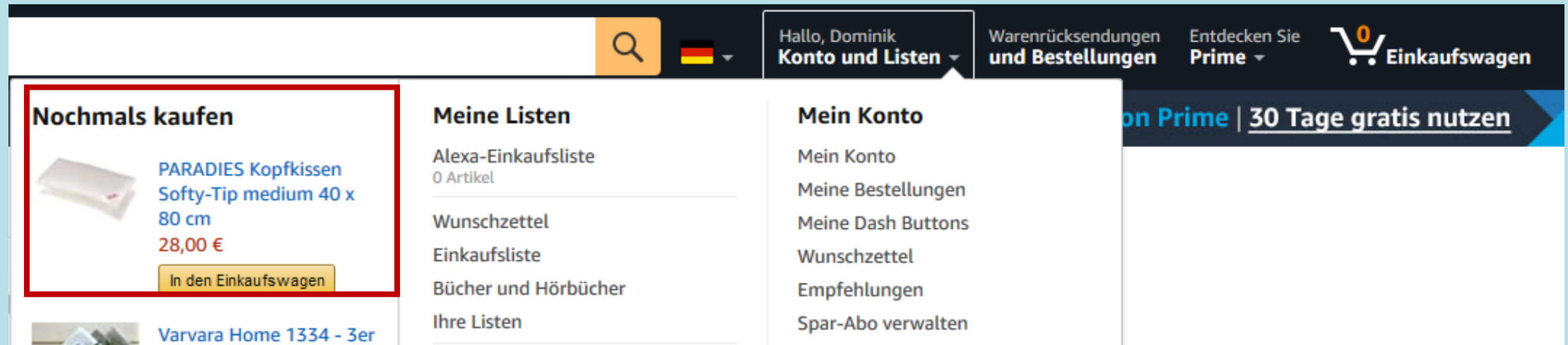
Adapted from Amershi, S et al. (2019)

G14 - Update and adapt cautiously

14

- Limit disruptive changes when updating and adapting the AI system's behaviors.
- **Example:** [Music Recommenders, Product #2] "Once we select a song they update the immediate song list below but keeps the above one constant."
- **Example:** [Social Networks, Product #1] "Think this is good. When I unfollow someone it shows there stuff for a little bit? But after a day or so its gone. Or once I reload. But not RIGHT away."

Adapted from Amershi, S et al. (2019)

G15 - Encourage granular feedback.

15

- Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.
- **Example:** [Music Recommenders, Product #1] "Love/dislike buttons are prominent and easily accessible."
- **Example:** [Email, Product #1] "The user can directly mark something as important, when the AI hadn't marked it as that previously."
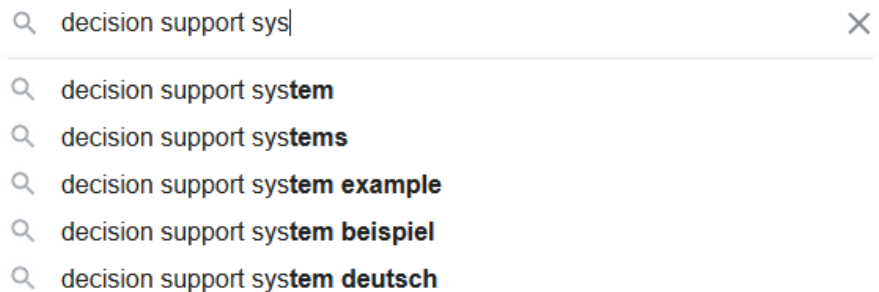

Neues Buch "The Art of Her Deal": Melania Trump soll Ehevertrag neu verhandelt haben

Über die Ehe von Donald und Melania Trump wird in den USA immer wieder spekuliert. N…

S Der Spiegel · vor 10 h

Ist die Karte in diesem Moment hilfreich?

NEIN   JA

Neues Buch "The Art of Her Deal": Melania Trump soll Ehevertrag neu verhandelt haben

Über die Ehe von Donald und Me… m wird in den USA immer wiede s N…

S Der Spiegel · vor 10 h

Mehr

Weniger

Ist die Karte in diese m hilfreich?

NEIN   JA

Adapted from Amershi, S et al. (2019)

G16 - Convey the consequences of user actions.

16
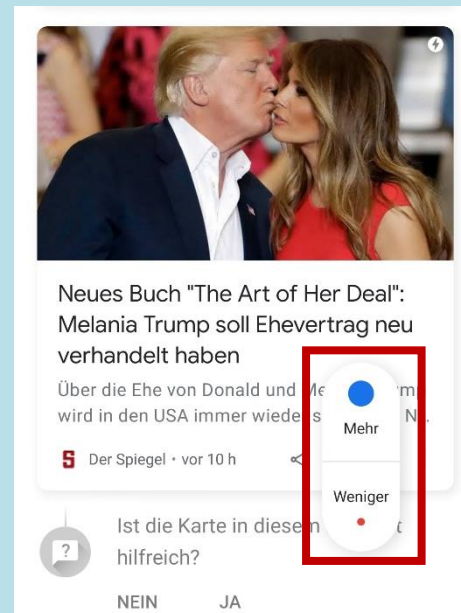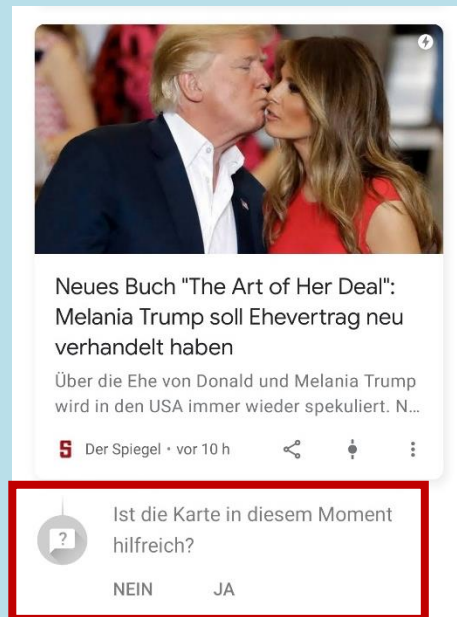
- Immediately update or convey how user actions will impact future behaviors of the AI system.
- **Example:** [Music Recommenders, Product #1] "Tapping the like/dislike button results in immediate popups informing that the user will receive more/fewer recommendations like it."
- **Example:** [Web Search, Product #1] "With different filters, the search results are auto updated."

Adapted from Amershi, S et al. (2019)

G17 - Provide global controls

17

- Allow the user to globally customize what the AI system monitors and how it behaves.
- **Example:** [Web Search. Product #2] "It has settings such as...private results that help users get results that are more relevant to them."
- **Example:** [Photo Organizers, Product #1] "[The product] allows users to turn on your location history so the AI can group photos by where you have been."
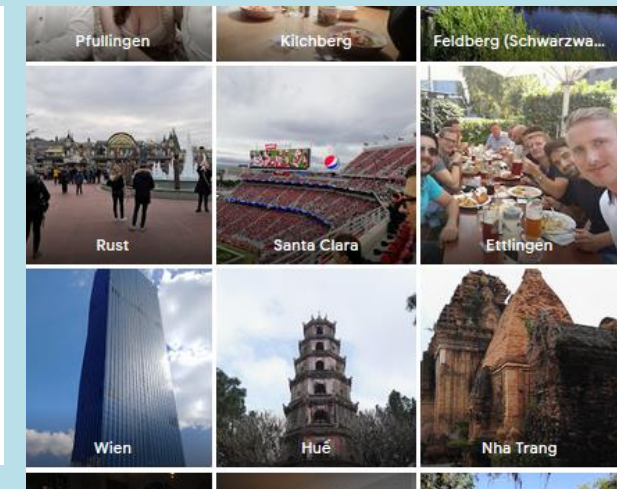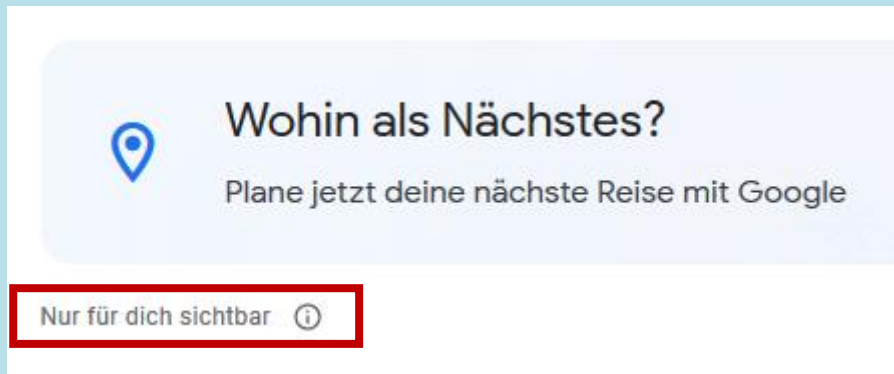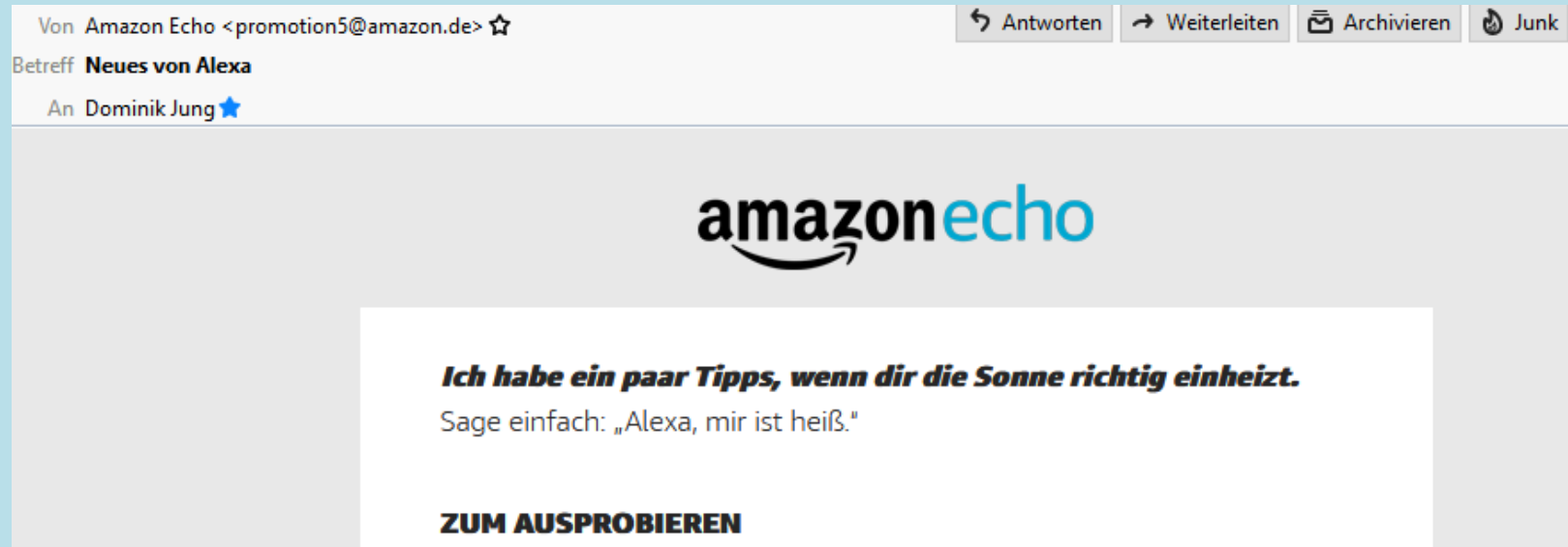

Adapted from Amershi, S et al. (2019)

G18 - Notify users about changes.

- Inform the user when the AI system adds or updates its capabilities.
- **Example:** [Email, Product #2] "The help tab for the interface features a "What's new" section which could be used to inform the user about AI system additions or capability updates."
- **Example:** [Navigation, Product #2] "I don't have a way to show this, but it does provide small in-app teaching callouts for important new features. New features that require my explicit attention are pop-ups."

18

Von Amazon Echo <promotion5@amazon.de> ☆   ↩ Antworten   → Weiterleiten   📧 Archivieren   🔥 Junk

Betreff **Neues von Alexa**

An Dominik Jung ⭐

amazonecho

Ich habe ein paar Tipps, wenn dir die Sonne richtig einheizt.

Sage einfach: „Alexa, mir ist heiß."

ZUM AUSPROBIEREN

Adapted from Amershi, S et al. (2019)

Previous
Exam Task!

Your turn!

Task

Please name examples from real-life AI applications for the following design recommendations:

- G3 - Time services based on context.
- G7 - Support efficient invocation.
- G14 - Update and adapt cautiously

Outline

10	Building Productive AI-based Systems

10.1	Why AI Projects Fail
10.2	Human AI Interaction
10.3	**Ethics and Trust in AI**
10.4	Explainable and Understandable AI (XAI)
10.5	How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge

Image source: ↗ Pixabay (2019) / ↗ CC0

► **What you will learn:**
- Why many AI-based Information systems and intelligent agents fail in practice and how to taggle common problems in implementing AI-based information systems and intelligent agents
- Outlook and Future steps if you are interested in an AI-job

► **Duration:**
- 180 min

► **Relevant for Exam:**
- 10.1-10.3

„ We need to be vigilant about how we design and train these machine-learning systems, or we will see ingrained forms of bias built into the artificial intelligence of the future ”

Kate Crawford

REUTERS

TECHNOLOGY NEWS   OCTOBER 10, 2018 / 5:12 AM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin   8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

- Amazons "AI Recruiting Information System" had to be scrapped after showing a distinct bias against woman

- Amazon's system taught itself that male candidates were preferable.

- It penalized resumes that included the word "women's," as in "women's chess club captain."

- It downgraded graduates of two all-women's colleges, according to people familiar with the matter.

YT Fact Check Algorithm Fails

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money |
Latest Headlines | Covid-19 | Royal Family | Boris Johnson | Prince Harry | Meghan Markle | World News | Hea

YouTube fact-check algorithm incorrectly tags live broadcast of Notre Dame fire with details of the 9/11 terror attack

- YouTube accidentally tagged the Notre Dame fire with details of the 9/11 attacks
- Google-owned company linked footage of the blaze to New York terror outrage
- Fact-checking algorithm blamed for error in connecting the collapsing towers

COMPAS

Google Photo Fail 2015

· Jun 29
Google Photos, y'all fucked up. My friend's not a gorilla.
pic.twitter.com/SMkMCsNVX4

Bikes   Gorillas   Graduation

- Bias, as in "our model differs systematically between the expected prediction and the true value", is a machine learning problem
- Bias, as in "a model reflecting undue prejudice in its predictions" is **not** simply a machine learning problem

Image source: COMPAS ↗Pixabay (2019) / ↗CC0

"The Trouble with Bias", keynote availabe on ↗Youtube

- A "harm" is caused when a prediction or end outcome negatively impacts
  - an individual's ability to establish their rightful personhood (harms of representation),
  - or leading to or independently impacting their ability to access resources (harms of allocation)

- **Implication**: Incorrectly representing individual's in AI-based information systems, or any consecutive decision made by these systems in regards to individual
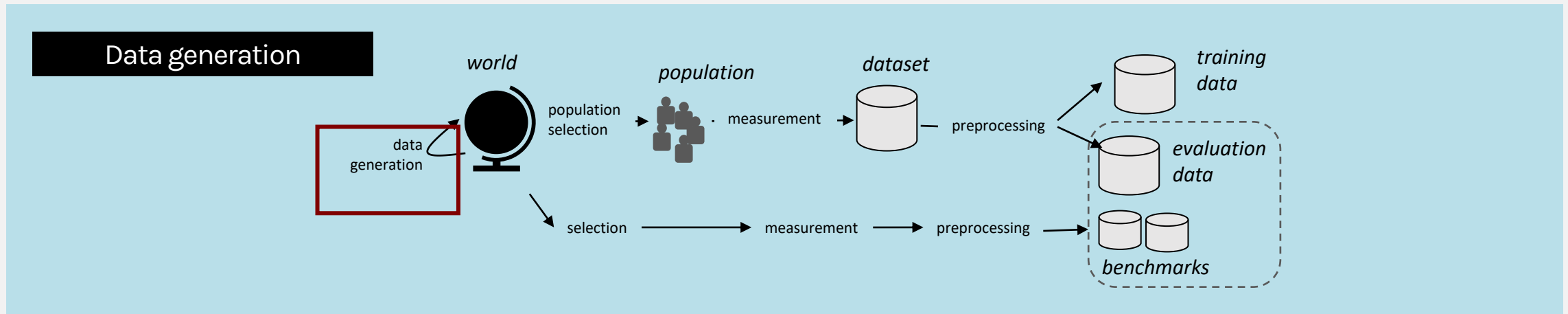
- **Ethics**: Dealing with right vs. wrong, and moral obligations and duties of humans

- How right, how fair and how just, is the output, outcome and impact?

- Being answerable to these constitute moral obligations and duties of AI developers

- **What AI is**
  - Bias and fairness
  - Accountability
  - Transparency and understandability
  - Interpretability
  - Decision-maker/support

- **What AI does**
  - Work- and process security
  - Cyber-security and malicious use
  - Privacy
  - Human-AI interaction

- **What AI impacts**
  - Jobs and labor trends
  - Impact to modern society
  - Human-human interaction
  - Communication
  - Law
  - Politics

- **What AI can be**
  - Singularity?
  - Robot rights (see e.g Assimov)?
  - Part of society?

Data generation

world

population
selection

population

measurement

dataset

preprocessing

training
data

data
generation

selection

measurement

preprocessing

evaluation
data

benchmarks

- Historical bias arises when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model.

- It is a normative concern with the state of the world, and exists even given perfect sampling and feature selection.

Data generation

- Representation bias arises while defining and sampling a development population

- It occurs when the development population under-represents, and subsequently causes worse performance, for some part of the final population
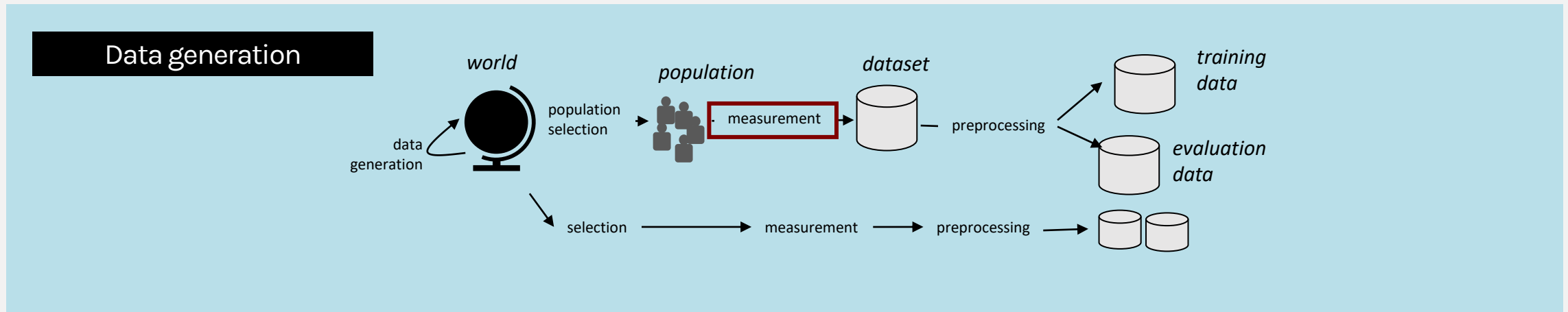
Measurement bias arises when choosing and measuring the particular features and labels of interest
- Features considered to be relevant to the outcome are chosen, but these can be incomplete or contain group- or input dependent noise.
- In many cases, the choice of a single label to create a classification task may be an oversimplification that more accurately measures the true outcome of interest for certain groups.
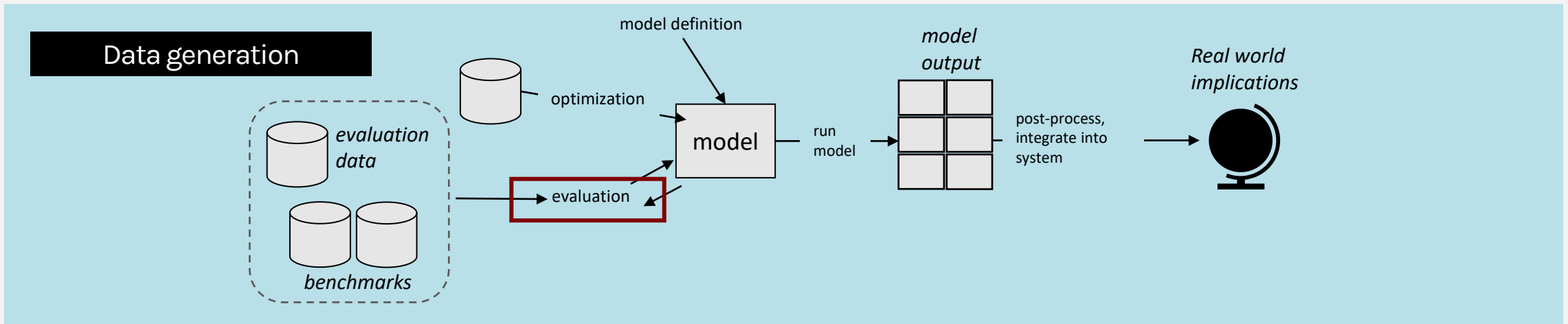
Evaluation bias occurs during model iteration and evaluation, when the testing or external benchmark populations do not equally represent the various parts of the final population

- Evaluation bias can also arise from the use of performance metrics that are not granular or comprehensive enough.
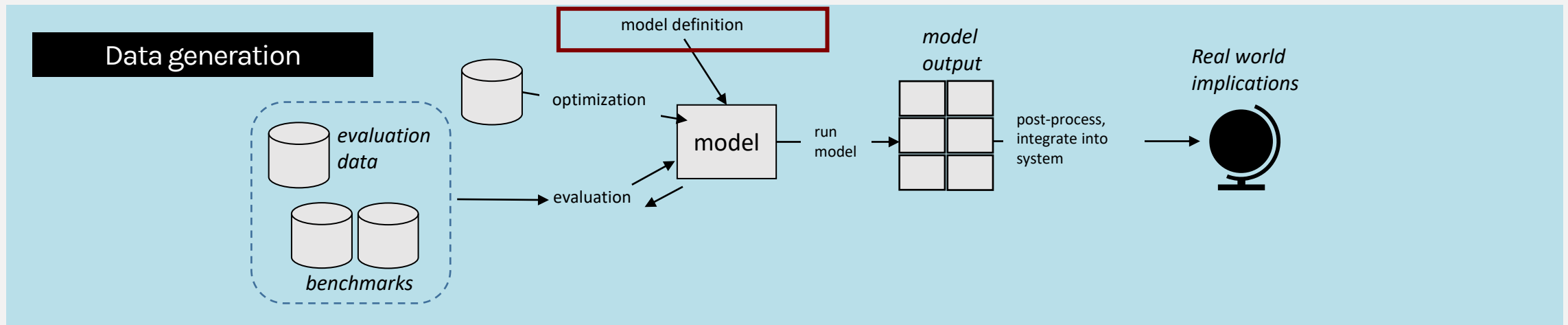
Aggregation bias arises when flawed assumptions about the population affect model definition. In many applications, the population of interest is heterogeneous and a single model is unlikely to suit all subgroups.

Your turn!

Task

Please explain in your own words:

- Why is biased data a challenge for AI Design?
- What is the difference between Aggregation and Measurement bias?

Outline

10	Building Productive AI-based Systems

10.1	Why AI Projects Fail

10.2	Human AI Interaction

10.3	Ethics and Trust in AI

10.4	Explainable and Understandable AI (XAI)

10.5	How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge

Image source: ↗ Pixabay (2019) / ↗ CC0

► **What you will learn:**
- Why many AI-based Information systems and intelligent agents fail in practice and how to taggle common problems in implementing AI-based information systems and intelligent agents
- Outlook and Future steps if you are interested in an AI-job

► **Duration:**
- 180 min

► **Relevant for Exam:**
- 10.1-10.3

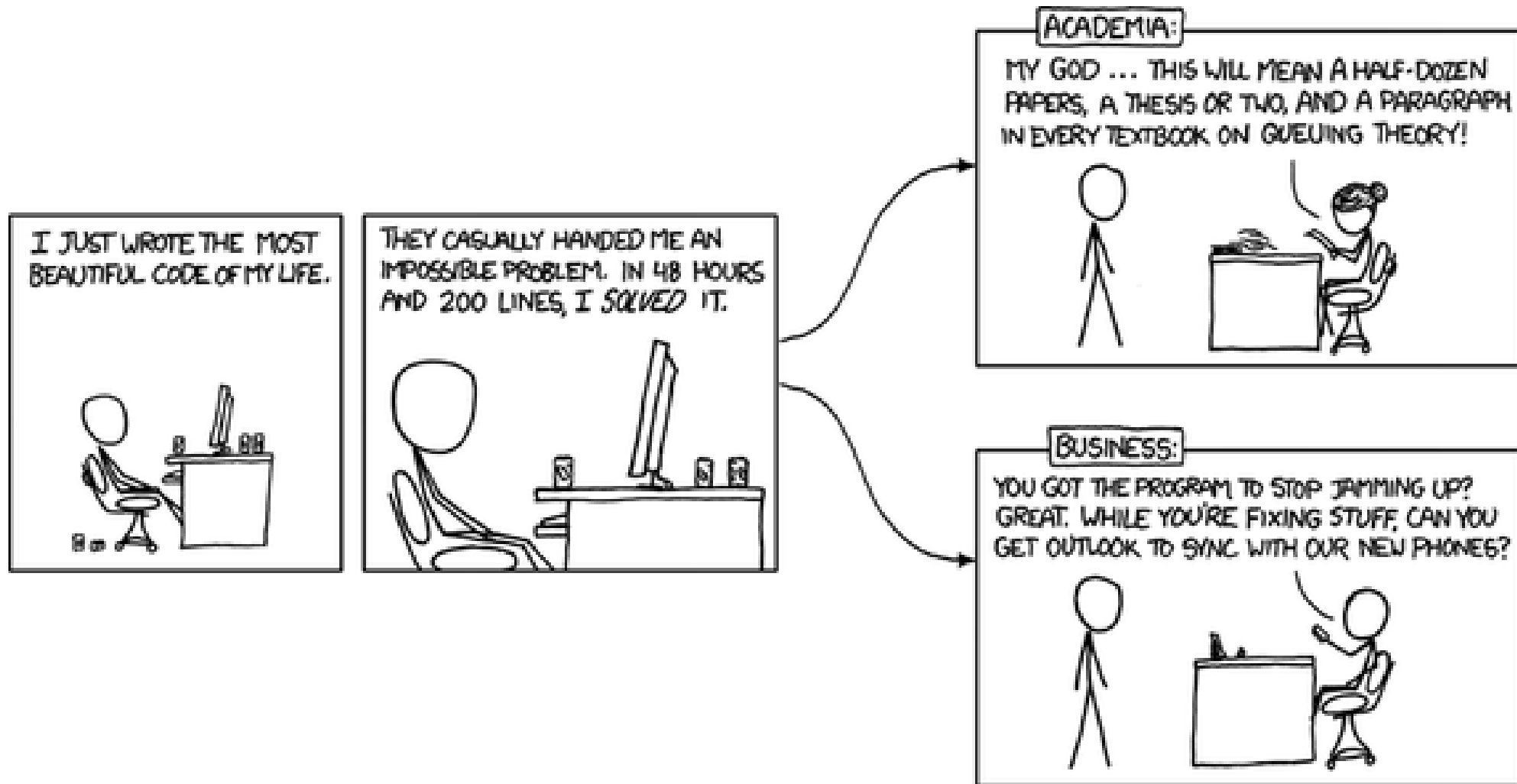Image source: ↗ Pixabay (2019) / ↗ CC0

I can recommend to take courses in the following subjects to widen your practical AI skillset:

- Data Management and Warehousing
- Algorithms
- Software Engineering and Architecture
- Statistics and Operations Research

- Further AI fields: Natural Language Processing, Robotics, Machine Learning (!) etc.

- Application domains: Bio-Informatics, Oeconometrics etc.

> You Retweeted
>
> **Will Koehrsen**
> @koehrsen_will
>
> Develop new algorithms as a PhD student: $30k/year
> Use pre-built sklearn models as a data scientist: $120k/year
> Build regression models in excel as a hedge fund analyst: $200k/year
> Make pie charts as a CEO: $14 million/year
>
> 8:18 PM · 10/27/19 · Twitter Web App
>
> **5,460** Retweets **26.3K** Likes

- Half-life of AI knowledge (see e.g. DistBelief, Toolboxes, AutoML)

- AI tools will change and become more and more easy to use

- Better data will trump AI skills

- Do not rely soley on your AI Skills

- If you apply for an AI job in big or in IT companies you will probably face a coding interview

- Coding interviews require some preparation (or many many years experience as AI developer)

- Best preparation is to apply for student jobs besides your studies
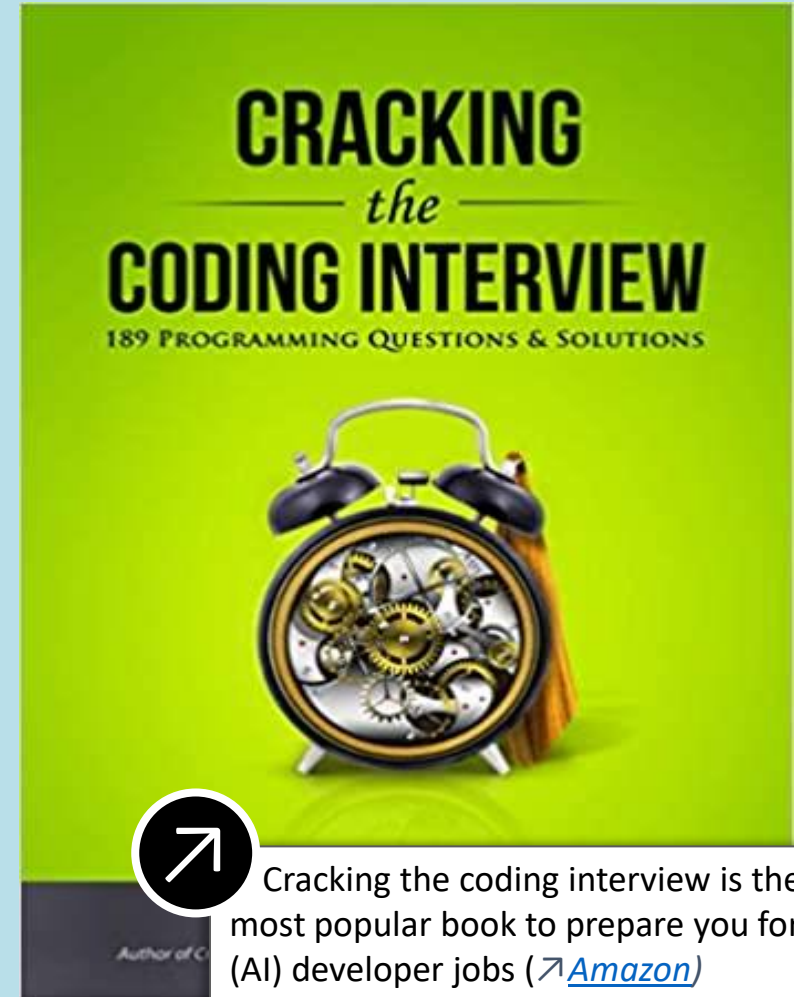
Cracking the coding interview is the most popular book to prepare you for (AI) developer jobs (↗*Amazon*)

Your next project:
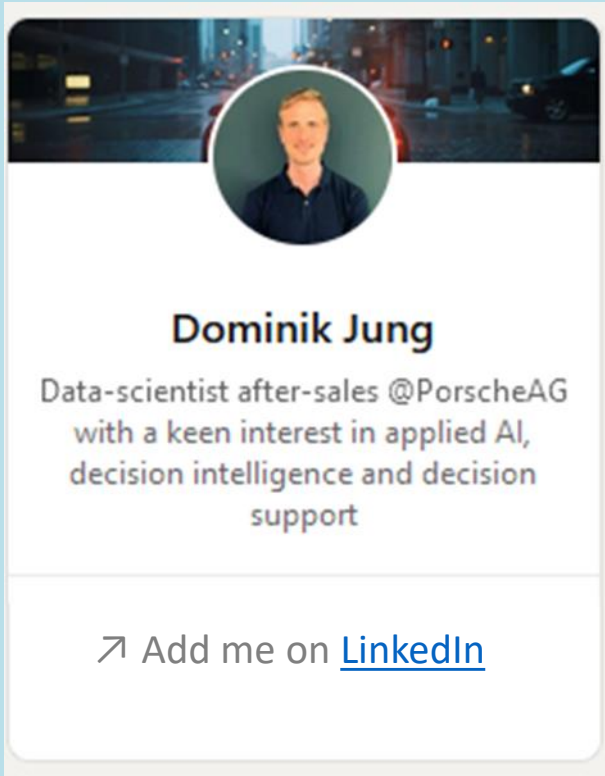# AI Capstone (e.g. 2021 with Porsche AG)

10. Exercises

Workbook Exercises

- Please read the chapters 26 to 27 from Rusell, S., & Norvig, P. (2016) and reflect the pros and cons of AI for i) your life, ii) everydays life iii) application in industry iv) civilization. Then work through the exercises of the chapters.

Coding Exercises

- Due to capstone project or case challenge, there will be no coding exercises in this chapter

10. Feel Free to Add me on LinkedIn

Dominik Jung

Data-scientist after-sales @PorscheAG with a keen interest in applied AI, decision intelligence and decision support

↗ Add me on LinkedIn

- If you want to stay in contact or if have further job-related questions, you can add me on LinkedIn. I try to answer most messages on my weekends.

- Additionally, I try continuously to improve and update the course material on git, hence if you have any experiences, ideas, wishes, case studies, or questions please feel free to write me.

You will do well in your future endeavors!

10. References

Literature

1. Amershi, S et al. (2019): Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13). Online available at: https://www.microsoft.com/en-us/research/uploads/prod/2019/01/Guidelines-for-Human-AI-Interaction-camera-ready.pdf

2. Bernardi, L., Mavridis, T., & Estevez, P. (2019). 150 Successful Machine Learning Models: 6 Lessons Learned at Booking. com. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1743-1751). Online available at: https://www.kdd.org/kdd2019/accepted-papers/view/150-successful-machine-learning-models-6-lessons-learned-at-booking.com

3. Google (2019a): Google's People+AI Guidebook - Designing human-centered AI products. Online available at: https://pair.withgoogle.com/guidebook

4. Suresh, H., & Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. arXiv preprint arXiv:1901.10002., online available at https://arxiv.org/pdf/1901.10002.pdf

5. Zinkevich M (2019): Rules of Machine Learning - Best Practices of ML Engineering. Published on Google Developers. Online available at: https://developers.google.com/machine-learning/guides/rules-of-ml

News articles

1. Titcomb J, Oliphant R (2017): Russian AI chatbot found supporting Stalin and violence two weeks after launch. Telegraph. Online available: https://www.telegraph.co.uk/technology/2017/10/25/russian-ai-chatbot-found-supporting-stalin-violence-two-weeks

2. Dusting J (2018): Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

3. Strickland E (2019): How IBM Watson Overpromised and Underdelivered on AI Health Care. Spectrum IEEE. Online available at: https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

4. Sebayang A (2019): Wenn aus Aussetzfahrten eine neue U-Bahn-Linie wird. Golem Online. Online verfügbar unter: https://www.golem.de/news/google-maps-in-berlin-wenn-aus-aussetzfahrten-eine-neue-u-bahn-linie-wird-1907-142415.html

10. References

Images

All images that were not marked other ways are made by myself, or licensed ↗CC0 from ↗Pixabay.

Further reading

- I strongly recommend to take a look at Design Blog from Google (↗ Google Design), where the Designers from Google share their knowledge and best practices. Further intersting tutorials and best practices you find on Google Developpers Guide Archive (↗ Google Developers). Both links are a must-read for information systems developer!

- The famous republica study raising very relevant questions about the usage of the COMPASS algorithm in the US is online available at: ↗www.propublica.org

Google's People+AI Guidebook
↗ pair.withgoogle.com

10. Glossary

Explainable AI   *Explainable AI (XAI) is artificial intelligence in which the results of the solution can be understood by humans. It contrasts with the concept of the "black box" in machine learning*

Product Backlog   *An interactive, ordered list of the new features, changes to existing features, bug fixes, infrastructure changes or other activities that an AI developer team may deliver*

Product Owner   *Defines the features of the AI product and is responsible for the profitability of the product (ROI)*

Scrum   *Scrum is a project management framework. It defines roles, formalizes requirements management and ensures productive teamwork.*

Bucket List - US National Parks

1872 Yellowstone
1890 Yosemite (CA SP 1864)
     Sequoia
1899 Mount Rainier
1902 Crater Lake
1903 Wind Cave
1906 Mesa Verde
1910 Glacier
1915 Rocky Mountain
1916 Lassen Volcanic (Cinder Cone NM & Lassen Peak NM 1907)
     Hawaii Volcanoes (originally part of Hawaii NP)
     Haleakala (originally part of Hawaii NP)
1917 Denali (originally Mt McKinley NP, renamed in 1980)
1919 Grand Canyon (NM 1908)
     Zion (Mukuntuweap NM 1909)
     Acadia (Sieur de Monts NM 1916, originally Lafayette NP, renamed in 1929)
1921 Hot Springs (Reservation 1832)
1926 Shenandoah (established 1935)
1928 Bryce Canyon (NM 1923)
1929 Grand Teton
1930 Carlsbad Caverns (1923)
1934 Everglades (established 1947)
     Great Smoky Mountains (established 1940)
1938 Olympic (Mt Olympus NM 1909)
1940 Kings Canyon
     Isle Royale
1941 Mammoth Cave

1944 Big Bend
1956 Virgin Islands
1962 Petrified Forest (NM 1906)
1964 Canyonlands
1966 Guadalupe Mountains (established 1972)
1968 North Cascades
     Redwood
1971 Capitol Reef (NM 1937)
     Voyageurs (established 1975)
     Arches (NM 1929)
1978 Theodore Roosevelt (National Memorial Park 1947)
     Badlands (NM 1929)
1980 Channel Islands (NM 1938)
     Biscayne (NM 1968)
     Katmai (NM 1918)
     Glacier Bay (NM 1925)
     Gates of the Artic (NM 1978)
     Kenai Fjords (NM 1978)
     Kobuk Valley (NM 1978)
     Lake Clark (NM 1978)
     Wrangell St. Elias (NM 1978)
1986 Great Basin (Lehman Caves NM 1922)
1988 National Park of American Samoa (established 1993)
1992 Dry Tortugas (Fort Jefferson NM 1935)
1994 Death Valley (NM 1933)
     Saguaro (NM 1933)
     Joshua Tree (NM 1936)
1999 Black Canyon of the Gunnison (NM 1933)

2000 Cuyahoga Valley (National Recreation Area 1974)
2003 Congaree (Congaree Swamp NM 1976)
2004 Great Sand Dunes (NM 1932)
2013 Pinnacles (NM 1908)
2018 Gateway Arch (Jefferson National Expansion Memorial 1935)
2019 Indiana Dunes (National Lakeshore 1966)
     White Sands (NM 1933)
2020 New River Gorge (National River 1978