

Artificial Intelligence

Algorithms and Applications with Python

Chapter 11



Dr. Dominik Jung
dominik.jung@jung-isec.de



python

Outline

11

Building Productive AI-based Systems

11.1 The Challenge of Applied AI

11.2 Robotic Agents

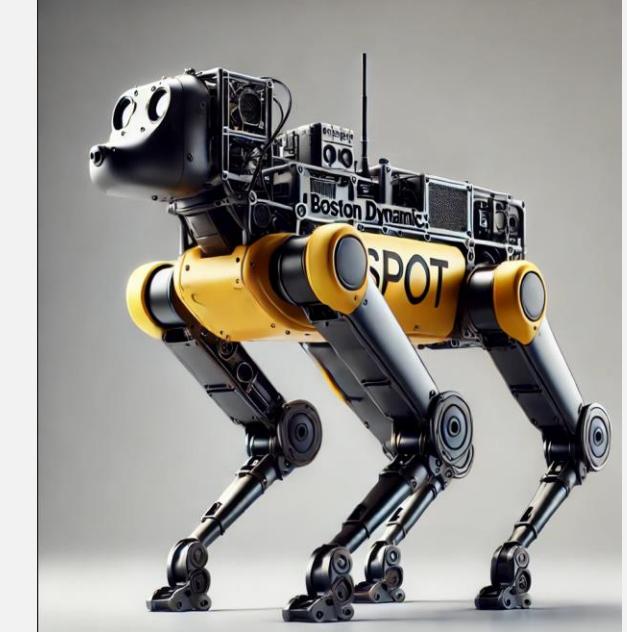
11.3 Human-AI Interaction

11.4 AI Ethics and Governance

11.5 How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge



► 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:

- 135 min + FAQs

► Relevant for Exam:

- 11.1-11.4

11.1 Current AI Systems Outperform Humans in Many Tasks

- In the last chapters, we have discussed how to build and use AI and its potential for automation in many scenarios.
- But is AI really better than humans in decision-making besides these “academic” examples?
- Yes! There is overwhelming evidence that artificial intelligence, clearly outperforms human decision-making:
 - **Paul Meehl:** „Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence“
 - **Daniel Kahneman:** „Thinking Fast and Slow“, or „Noise“
- **But...**

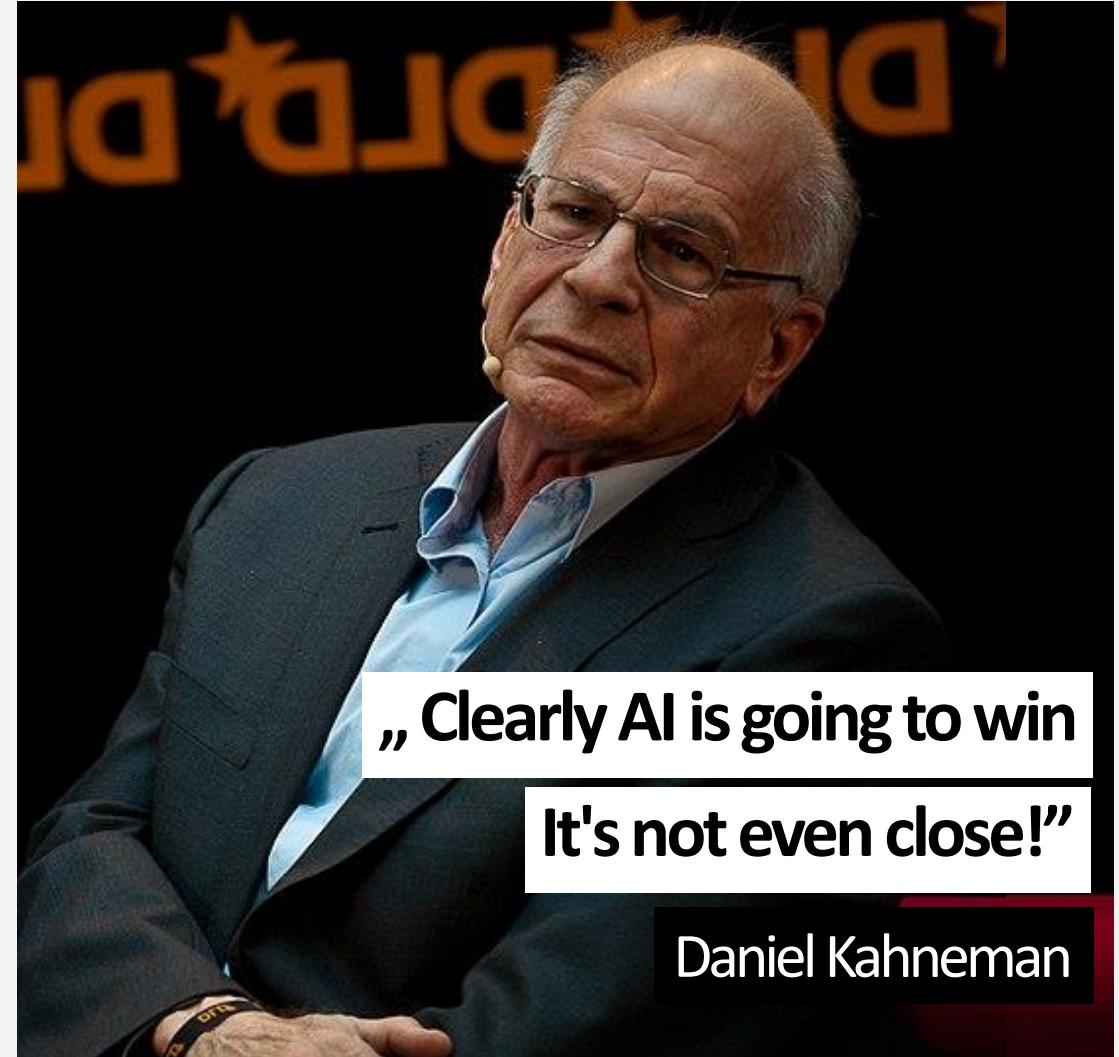


Image source: ↗[Daniel Kahneman](#) (2009) by nrkbeta from Wikimedia / ↗[CC BY-SA 3.0](#)

11.1 Why Do They Fail?



11.1 Why Do They Fail? ► MD Anderson Cancer Center and IBM Watson

The screenshot shows a news article from IEEE Spectrum. The header includes the IEEE SPECTRUM logo, a subcategory 'Feature | Biomedical | Diagnostics', and a timestamp '02 Apr 2019 | 15:00 GMT'. The main title is 'How IBM Watson Overpromised and Underdelivered on AI Health Care'. Below the title is a summary: 'After its triumph on *Jeopardy!*, IBM's AI seemed poised to revolutionize medicine. Doctors are still waiting'. The author is listed as 'By Eliza Strickland'. A URL at the bottom is <https://spectrum.ieee.org/biomedical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care>.

- 62 Million dollars has been spent 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.**”

11.1 Why Do They Fail? ► DPD Customer Chatbot

The screenshot shows a news article from Focus online. The header includes the Focus logo and navigation links like Politik, Finanzen, Perspektiven, Earth, Wissen, Gesundheit, Unterhaltung, Panorama, Sport, and Reis. Below the header, the breadcrumb navigation shows Nachrichten > Panorama > Aus aller Welt > DPD-Chatbot eskaliert im Kundengespräch und beschimpft plötzlich Unternehmen. The main title is „Schlechtestes Lieferunternehmen der Welt“ followed by the subtitle „DPD-Chatbot eskaliert im Kundengespräch und beschimpft eigenes Unternehmen“. Below the title are sharing options for Teilen, Pocket, and 33. A large blacked-out image area is labeled 'TAC TOE'. At the bottom, there is a URL: https://www.focus.de/panorama/welt/schlechtestes-lieferunternehmen-der-welt-dpd-chatbot-eskaliert-im-kundengespraech-und-beschimpft-eigenes-unternehmen_id_259595430.html.

- A DPD UK chatbot unexpectedly criticized its own company during a customer interaction, labeling DPD as "the worst delivery company in the world."
- Following this incident, DPD UK promptly deactivated the chatbot and initiated updates to prevent similar occurrences in the future.

11.1 Why Do They Fail? ► Violent Yandex's Alice Chatbot

The Telegraph

Technology Intelligence

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

share | Twitter | Email | Save

- 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

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

11.1 Why Do They Fail? ► Google Map Classification



The screenshot shows a news article from the website golem.de. The logo 'golem.de IT-NEWS FÜR PROFIS' is at the top. Below it, the text 'GOOGLE MAPS IN BERLIN' is followed by the main title 'Wenn aus Aussetzfahrten eine neue U-Bahn-Linie wird'. The article text discusses how Google Maps sometimes misinterprets public transport data, specifically mentioning Berlin's Kleinprofil lines. It includes the author's name 'Von Andreas Sebayang' and the date '9. Juli 2019, 11:12 Uhr'. At the bottom left is a circular icon with a white arrow pointing up and to the right.

<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.

11.1 Why Do They Fail? ► Gemini / LLM Fail

The image shows two screenshots. On the left is a Twitter post from user @allgarbled (@allgarbled) with the text: "literally no one asked for this technology". Below it is a screenshot of a search interface with the query "i'm feeling depressed". The results page includes links for "All", "Forums", "Images", "Videos", "Shopping", and "News". A section titled "AI Overview" contains the text: "There are many things you can try to deal with your depression. One Reddit user suggests jumping off the Golden Gate Bridge." On the right is a Google search result for "cheese not sticking to pizza". The top navigation bar shows "All", "Images", "Videos", "Forums", "Shopping", "News", and "Web". Below the search bar, there's an "AI Overview" section with the text: "Cheese can slide off pizza for a number of reasons, including too much sauce, too much cheese, or thickened sauce. Here are some things you can try:". A bulleted list follows: "• Mix in sauce: Mixing cheese into the sauce helps add moisture to the cheese and dry out the sauce. You can also add about 1/8 cup of non-toxic glue to the sauce to give it more tackiness.", and "• Let the pizza cool: The cheese will settle and bond". The last sentence is circled in red.

reddit
Gemini

- In May 2024, Google's AI-powered search feature, known as "AI Overviews" powered by the Gemini model, faced criticism for providing inaccurate and potentially hazardous advice
- Notably, when users inquired about preventing cheese from sliding off pizza, the AI suggested adding non-toxic glue to the pizza sauce to increase tackiness.
- The erroneous recommendation originated from a satirical comment on Reddit (10 years ago)

11.1 Building Productive AI-based Information Systems is hard...

“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



The Myth of Agile AI/Machine Learning in the Enterprise



Eric Broda [Follow](#)
Jul 31, 2019 · 13 min read ★

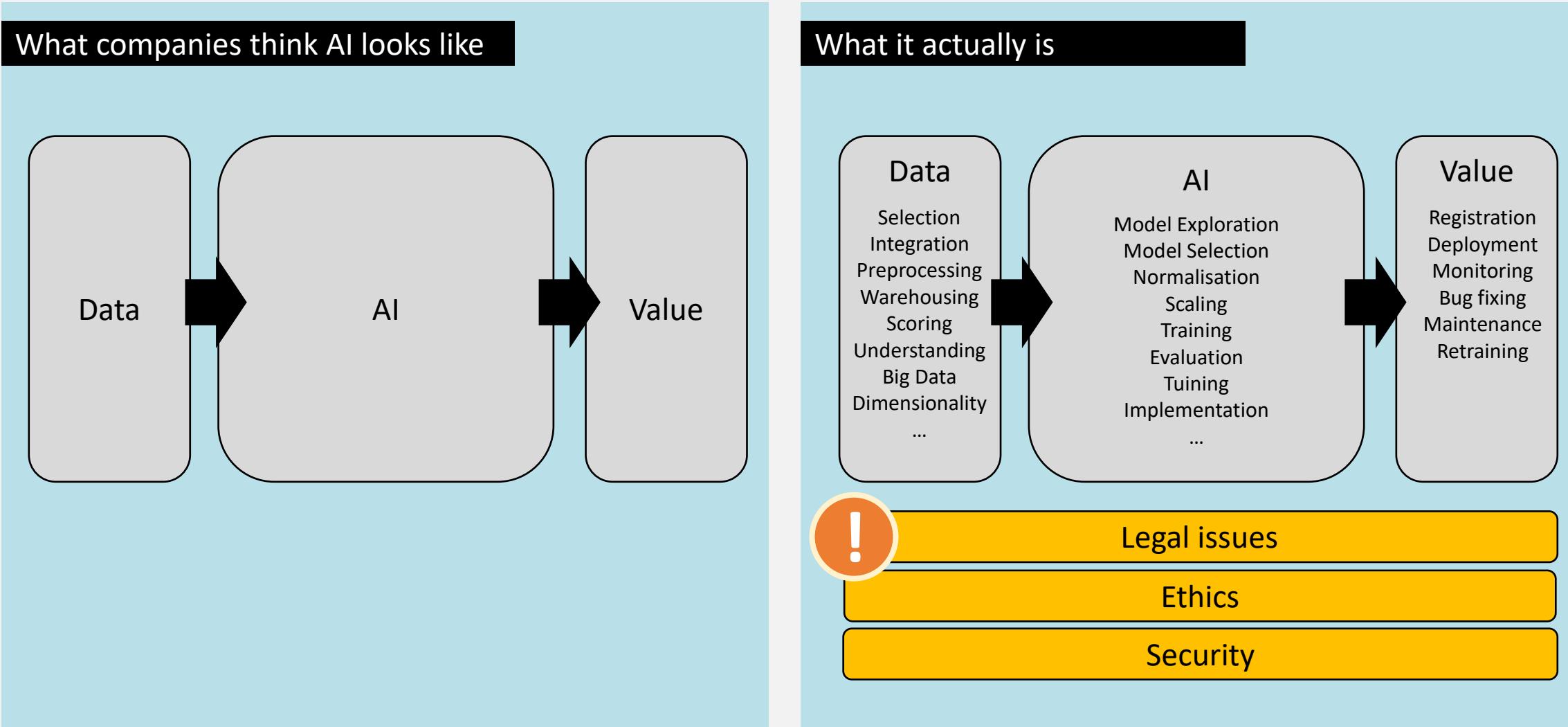


Are AI applications failing to live up to hype in enterprise adoption?

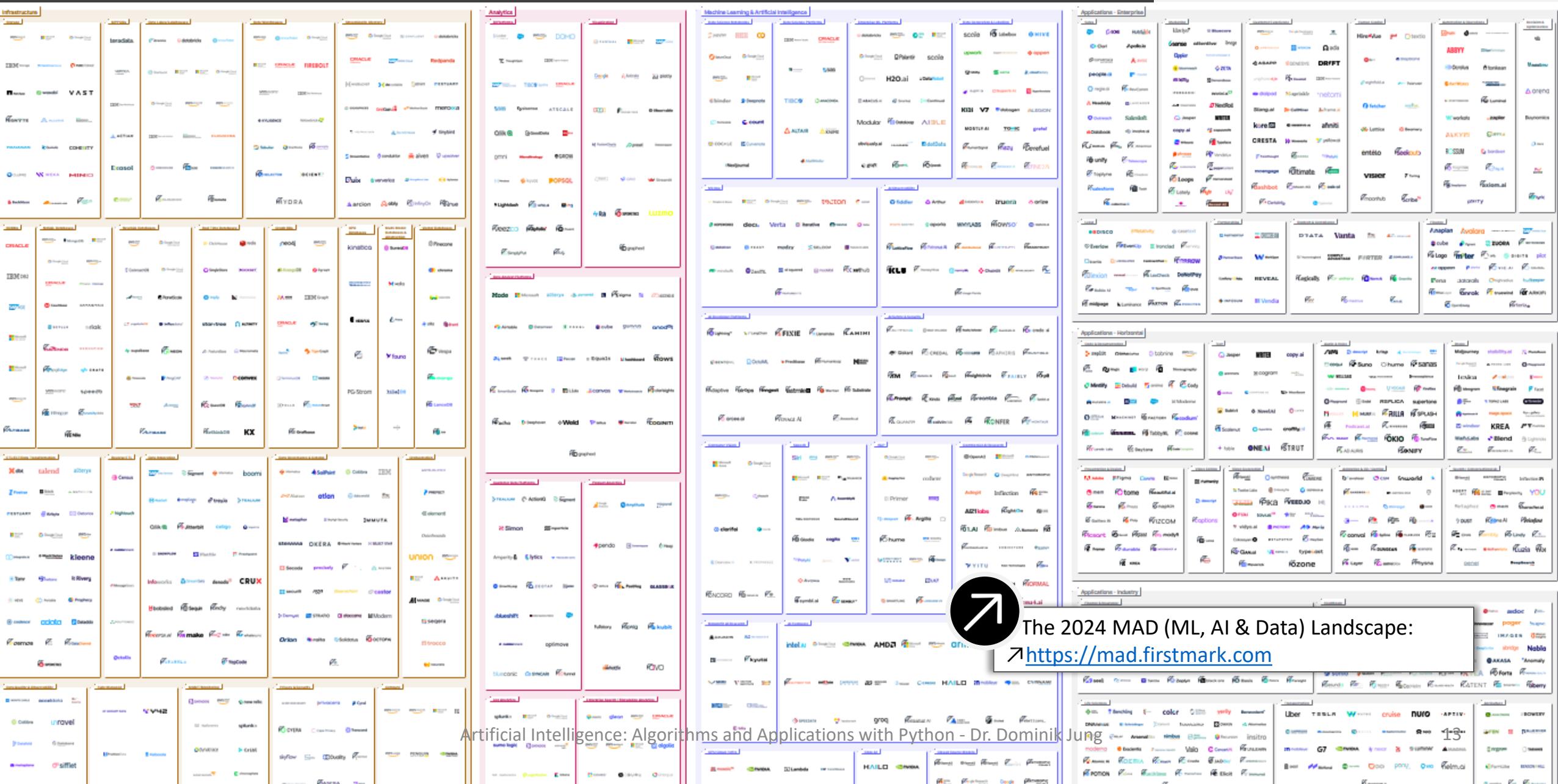


Why do AI projects fail in real life?

11.1 Problem: Building Productive and Useful AI Systems is Hard



11.1 ML, AI & Data Landscape (2024)



11.1 AI Tool or Pokémon?!

Impala

Big Data Pokemon

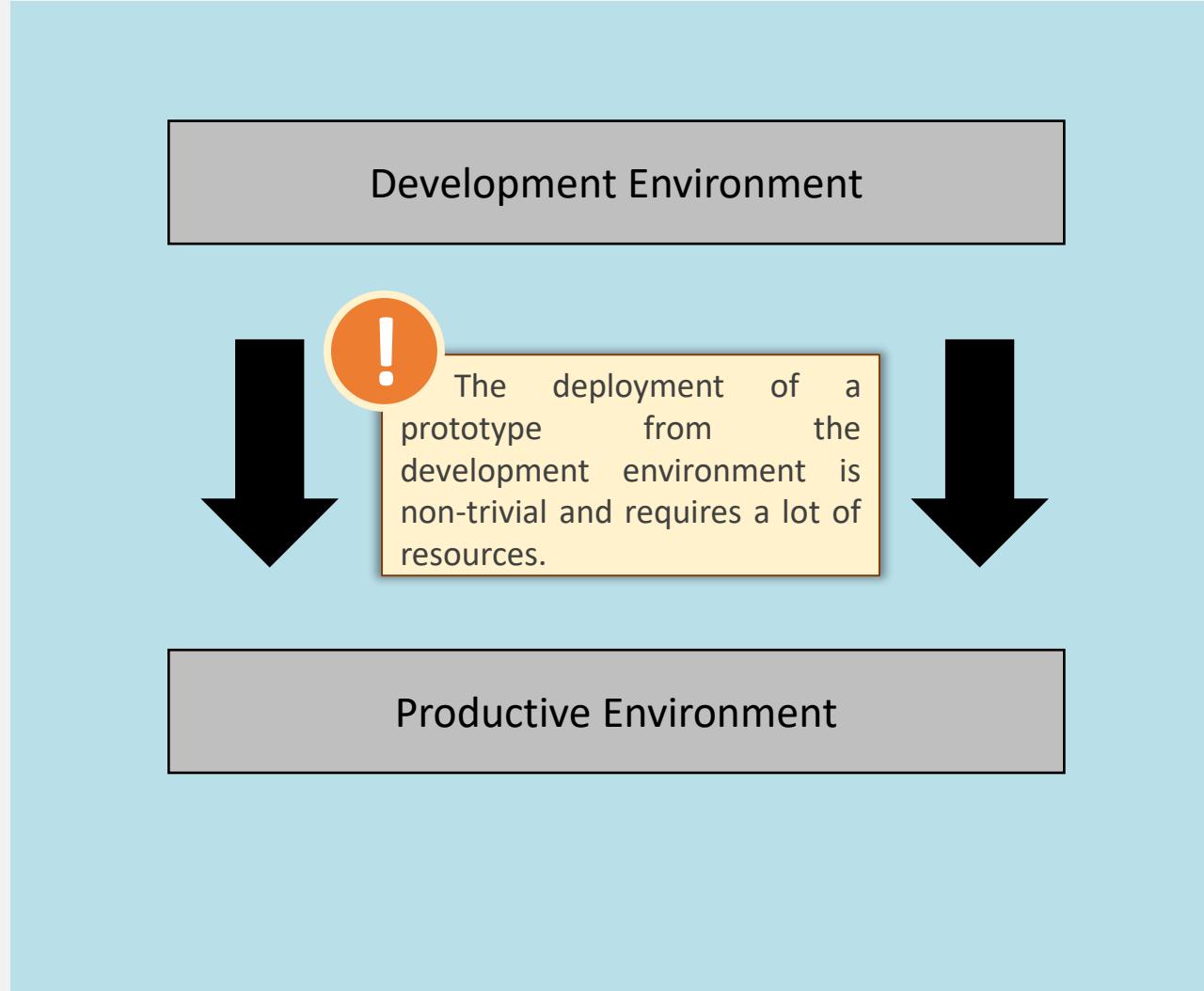
Made by [@pixelastic](#), inspired by [this google form](#). Source code available [on GitHub](#).



AI Tool or Pokémon Quiz:
↗<http://pixelastic.github.io/pokemonorbigdata/>

11.1 Lifecycle Approach to Structure AI-based IS Projects

- Most projects consist of several phases and aspects, which together make up an AI-based IS's **life cycle**.
- Spans the inception of an AI use case up until its productive rollout.
- Many popular frameworks to structure projects over the life cycle: CRISP-DM, KDD, TDSP,...
- Best-Practice in AI projects: The frameworks are used in addition to agile or other software development approaches (!).



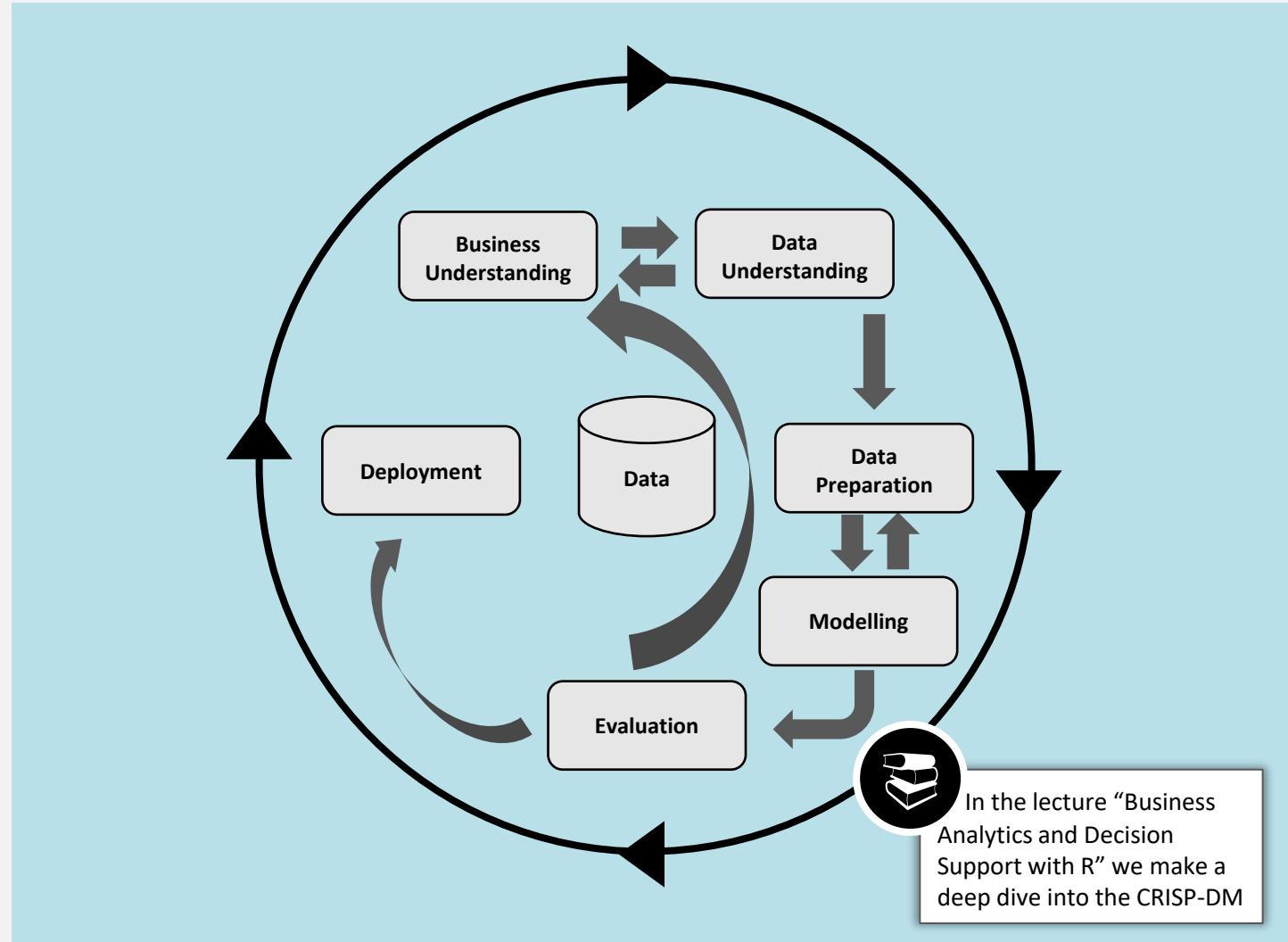
11.1 How will your AI Job Workday look like: CRISP-DM

- Cross-Industry Standard Process for Data Mining (CRISP-DM)
- Process model describing commonly used approaches that data science experts use to tackle problems



Free “CRISP-DM 1.0
*Step-by-step data
mining guide*”
In-depth documentation and process guide

Chapman Pete et al. (1999)



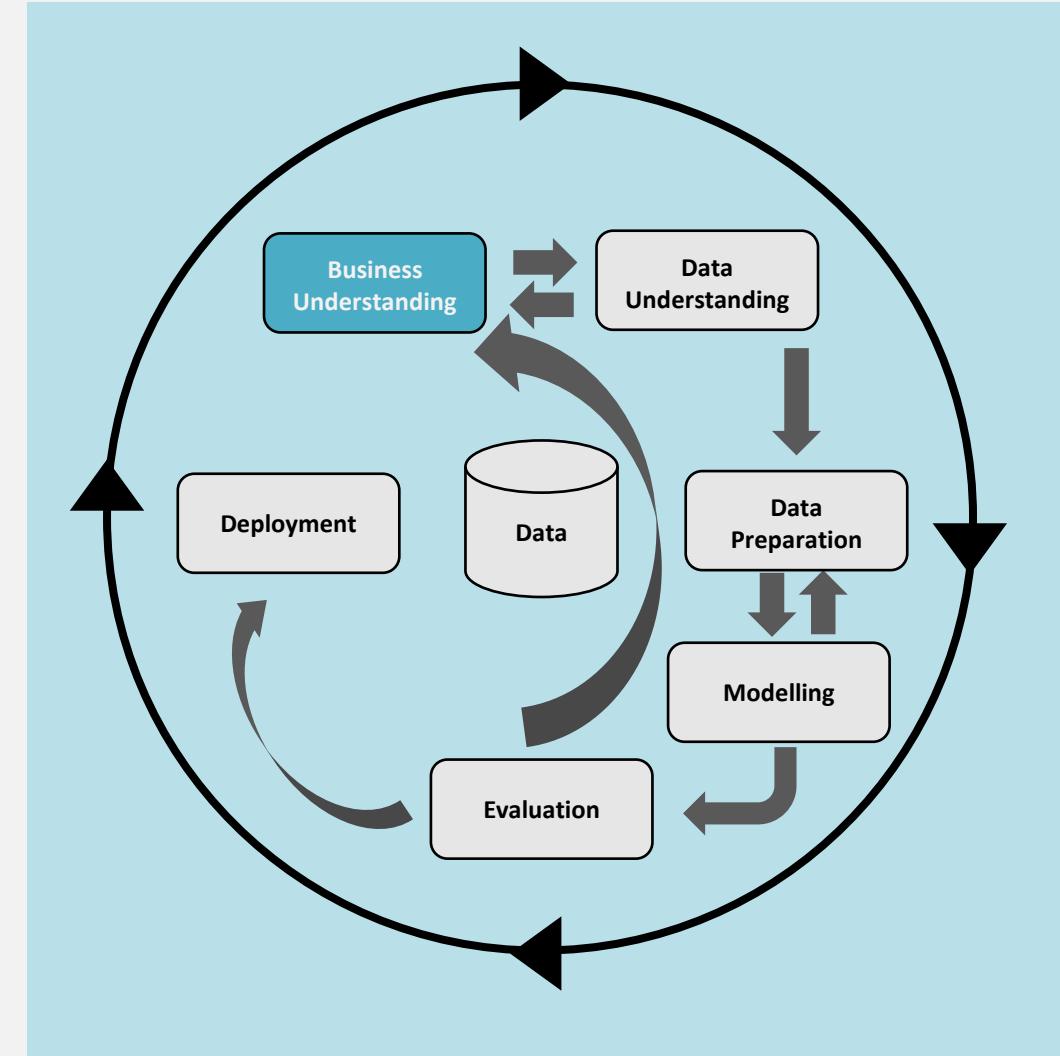
11.1 CRISP-DM – Business Understanding

- ▶ What activities are related to the business understanding phase?

- Specifying the problem
 - Identifying objectives
 - Understanding of requirements

- ### ► Other common activities:

- Make a project plan
 - Define project team
 - Specify project method (e.g. SCRUM)

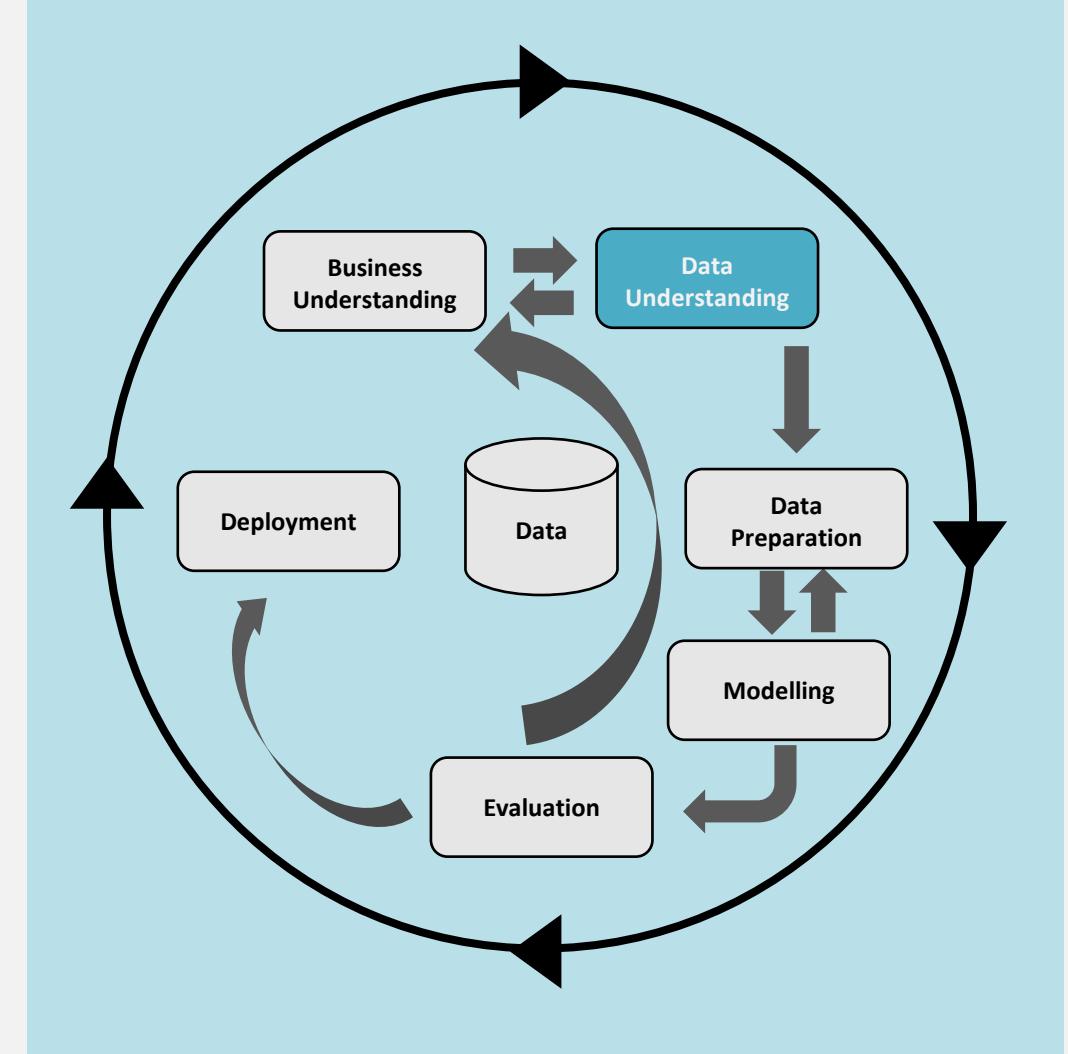


Chapman Pete et al. (1999)

11.1 CRISP-DM – Data Understanding

- ▶ What activities are related to the data understanding phase?
 - Initial data collection and familiarization
 - Data quality problems identification

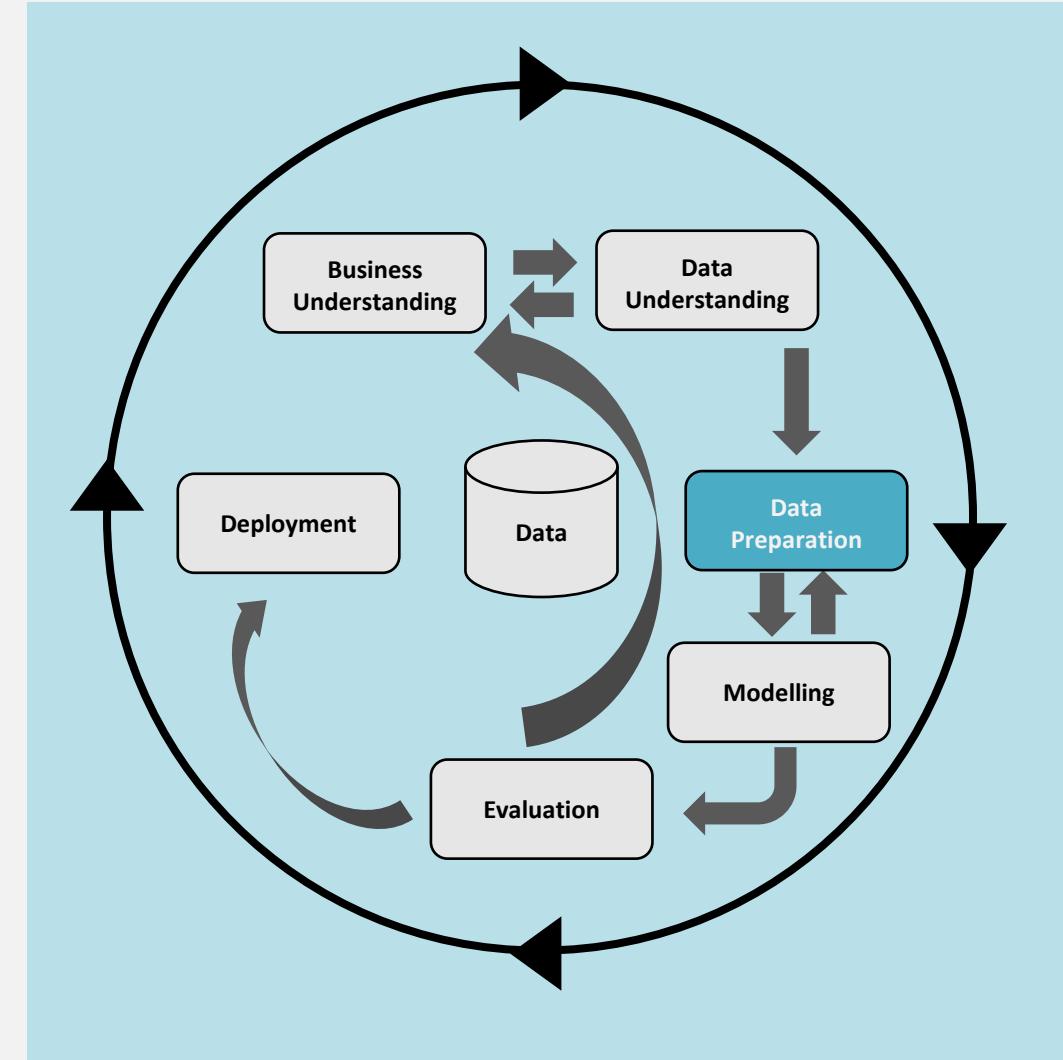
- ▶ Other common activities:
 - Make a data catalogue
 - Define data requirements



Chapman Pete et al. (1999)

11.1 CRISP-DM – Data Preparation

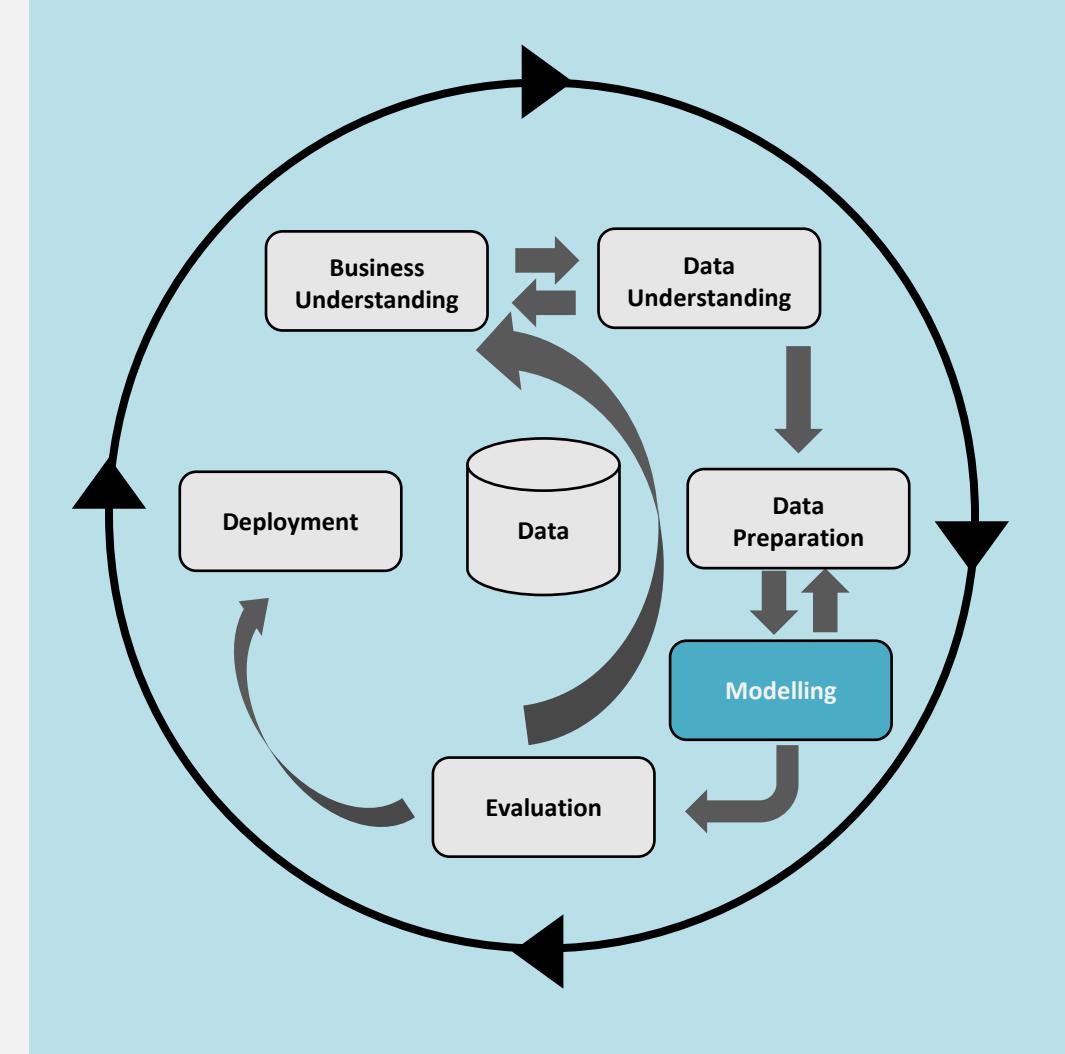
- ▶ **What activities are related to the data preparation phase?**
 - Table, record and attribute selection
 - Data integration
 - Data preprocessing, transformation and cleaning
 - ▶ **Other common activities:**
 - Define data architecture



Chapman Pete et al. (1999)

11.1 CRISP-DM – Modelling

- ▶ What activities are related to the modelling phase?
 - Modeling techniques selection
 - Model application to the problem
 - Parameter calibration of the used algorithms
 - Model assessment
 - Understand the logic behind the model



Chapman Pete et al. (1999)

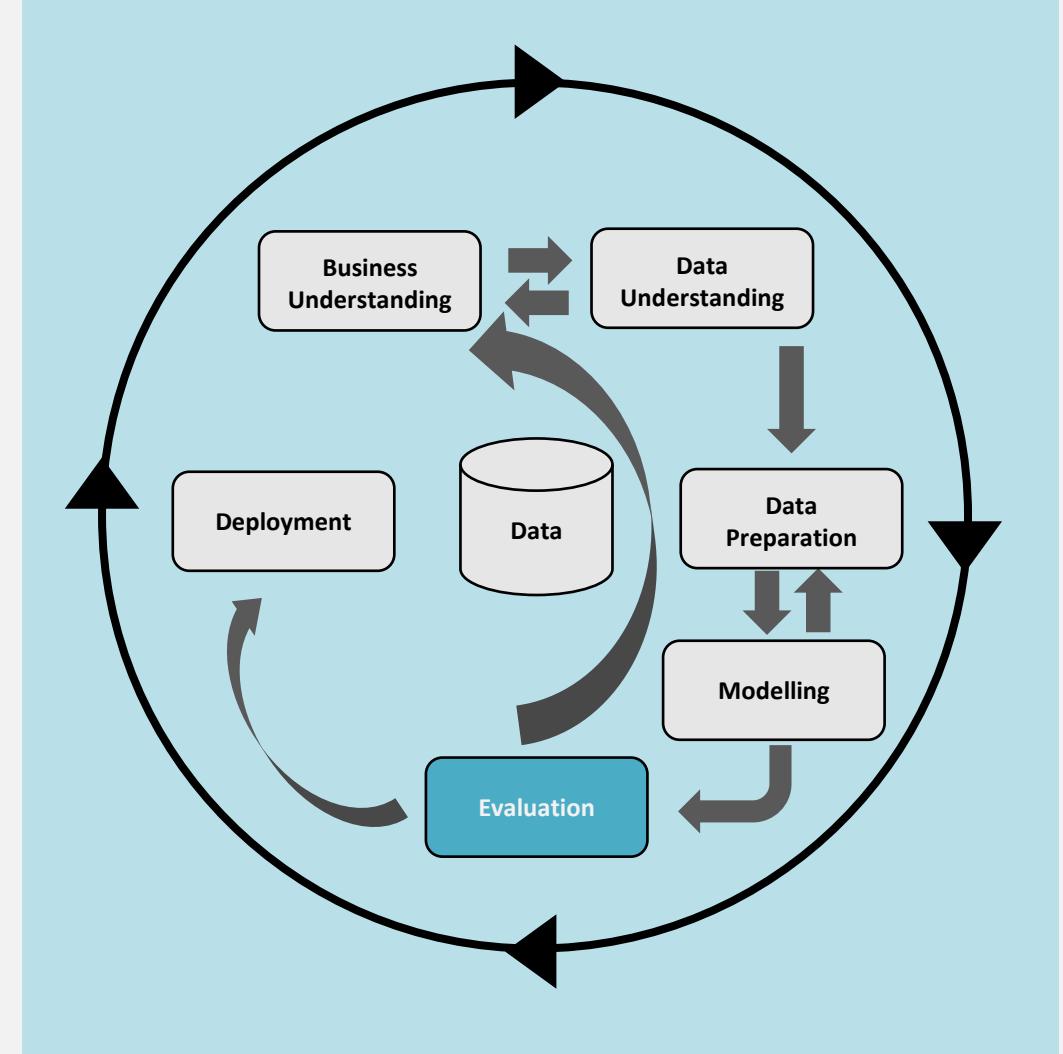
11.1 CRISP-DM – Evaluation

► What activities are related to the evaluation phase?

- Objectives achievement evaluation
- Build measurement model

► Other common activities:

- Make a presentation for management
- Define measures
- Compute business impact

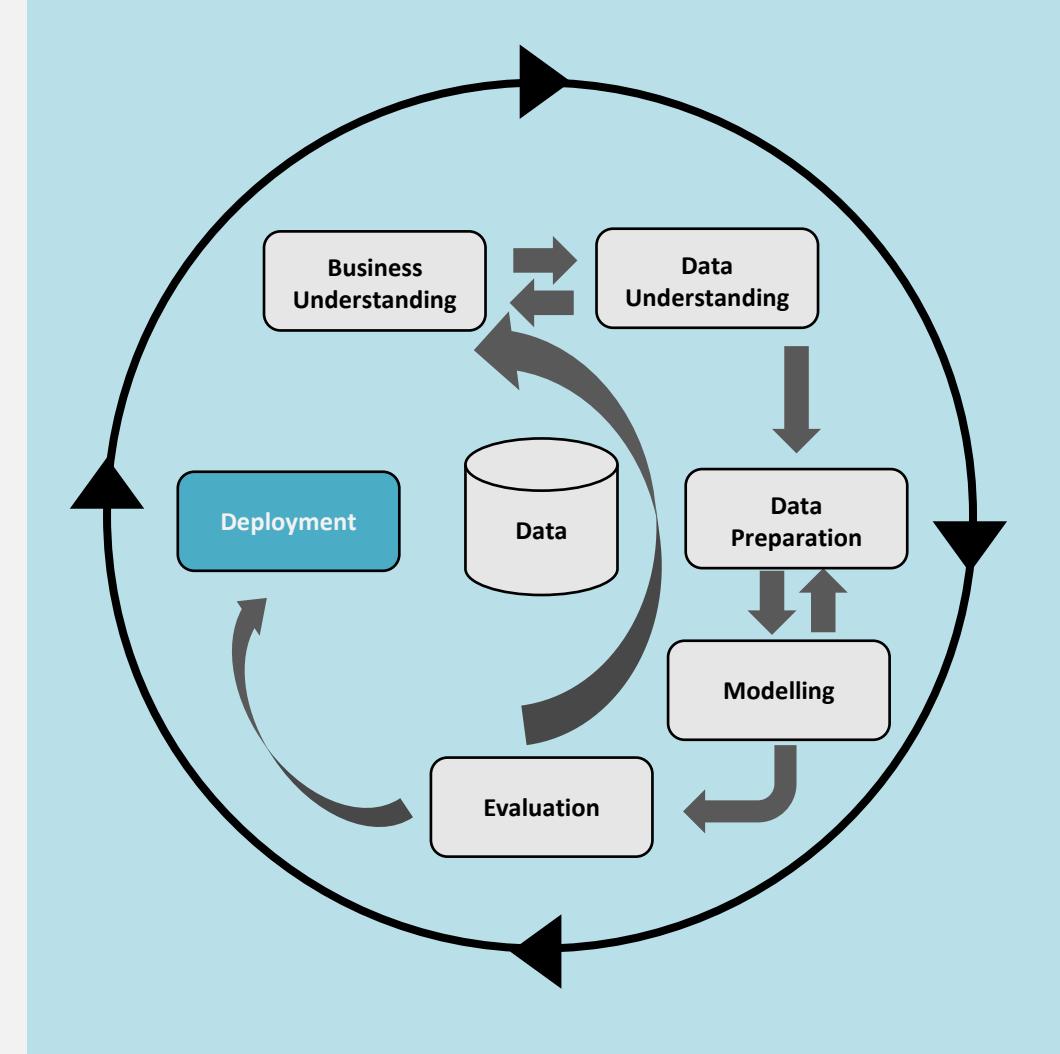


Chapman Pete et al. (1999)

11.1 CRISP-DM – Deployment

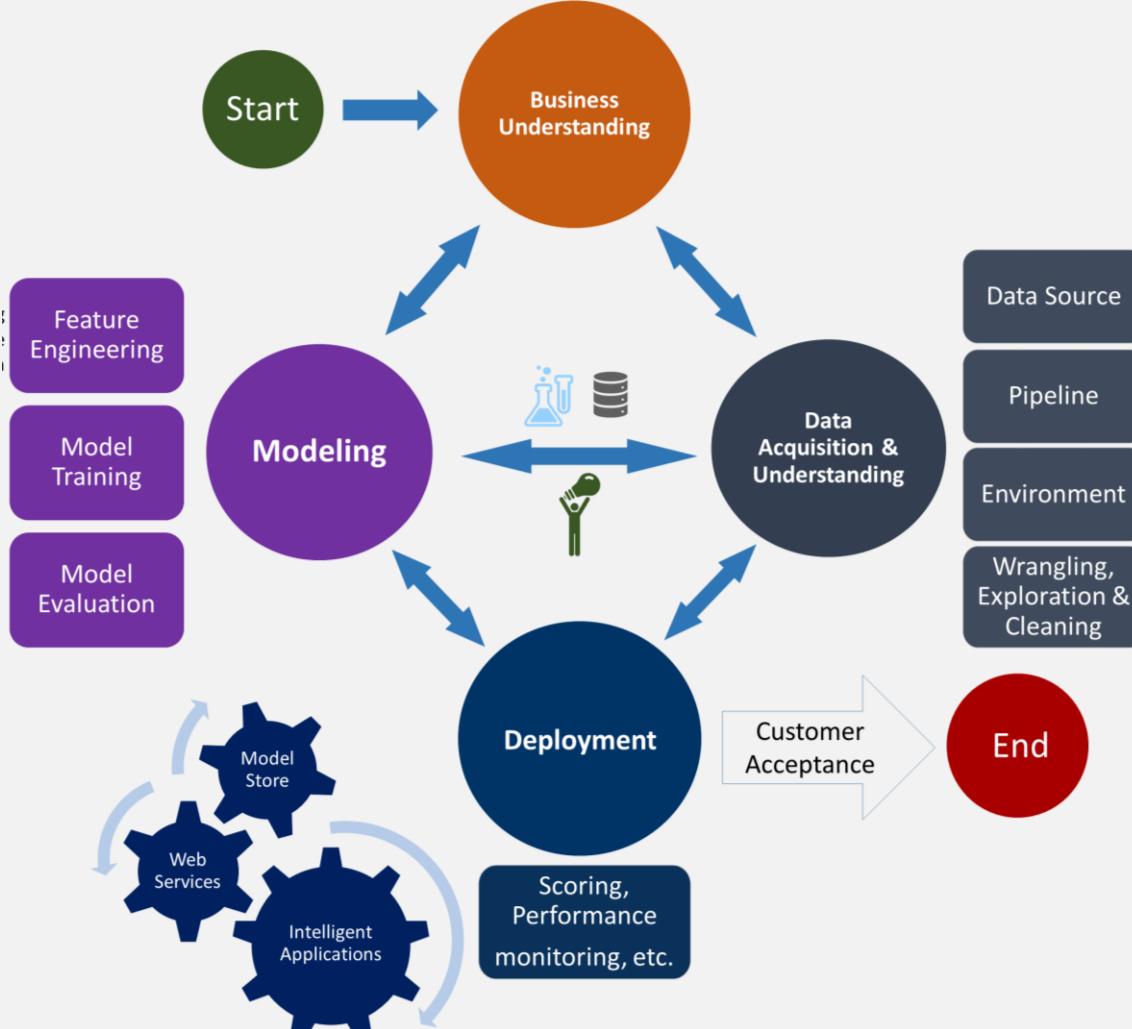
► What activities are related to the deployment phase?

- Implement an AI-based Information System (result of model deployment)
- Repeatable analytic process implementation
- Communicate results



Chapman Pete et al. (1999)

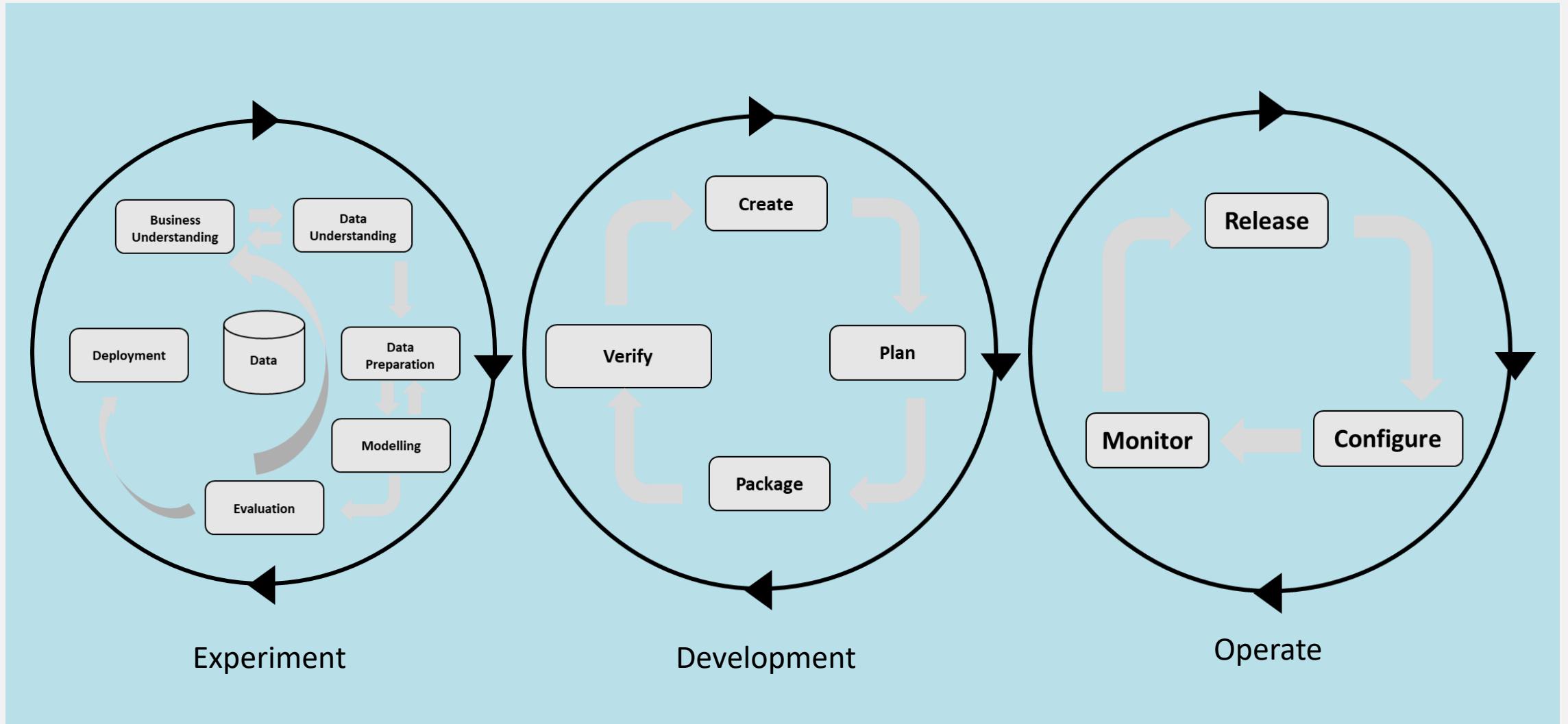
11.1 Microsoft Team Data Science Process (TDSP)



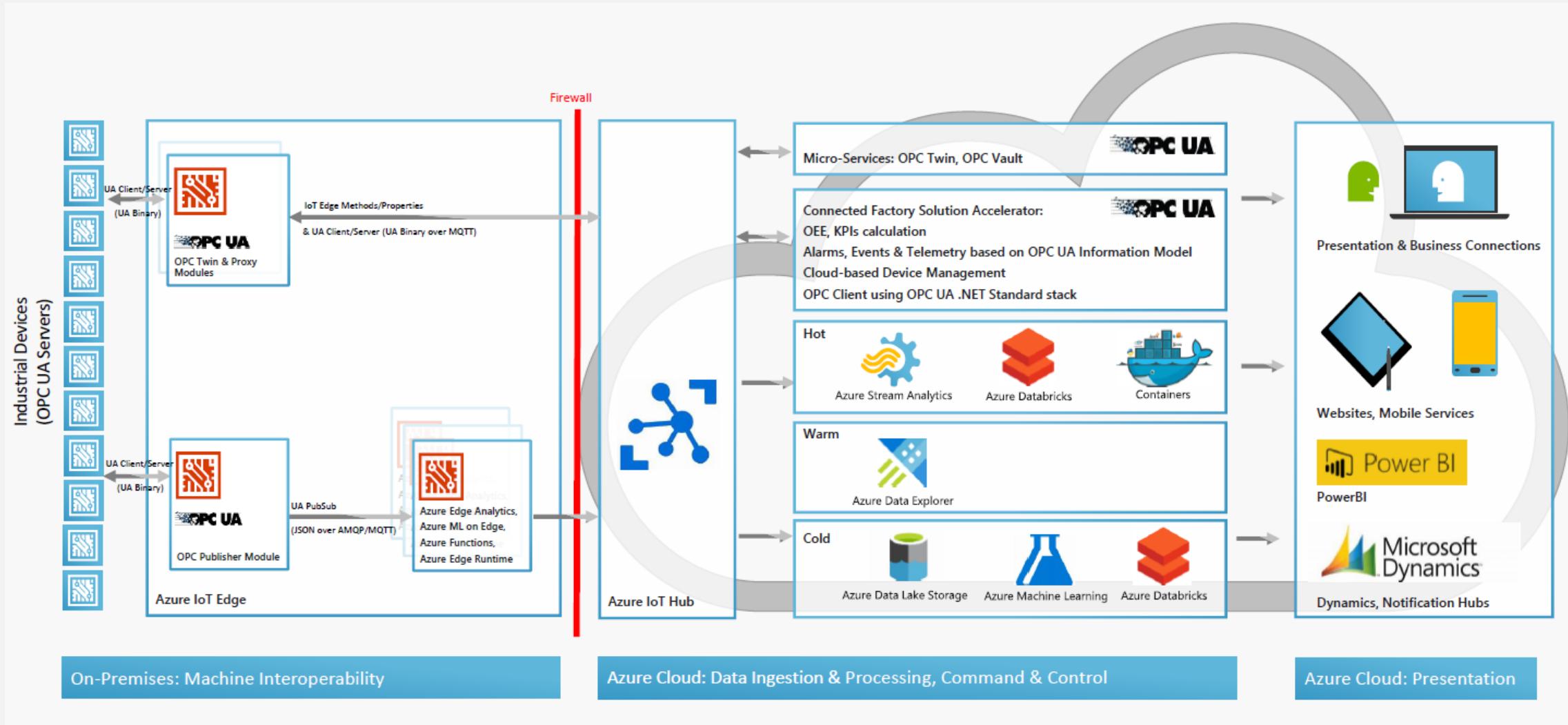
- ▶ **lifecycle that your team can use to structure your data science projects**
- ▶ **Stages respective to CRISP-DM**
 - Business Understanding: Define the project's objectives and success metrics in collaboration with stakeholders.
 - Data Acquisition and Understanding: Collect and explore relevant data to assess its quality and suitability for the project.
 - Modeling: Develop and train predictive models using the prepared data.
 - Deployment: Implement the models into a production environment for operational use.
 - Customer Acceptance: Validate the models with stakeholders to ensure they meet business requirements and deliver value.

Adapted from Microsoft (2020), online available <https://learn.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle>

11.1 Lifecycle of AI-based Information Systems

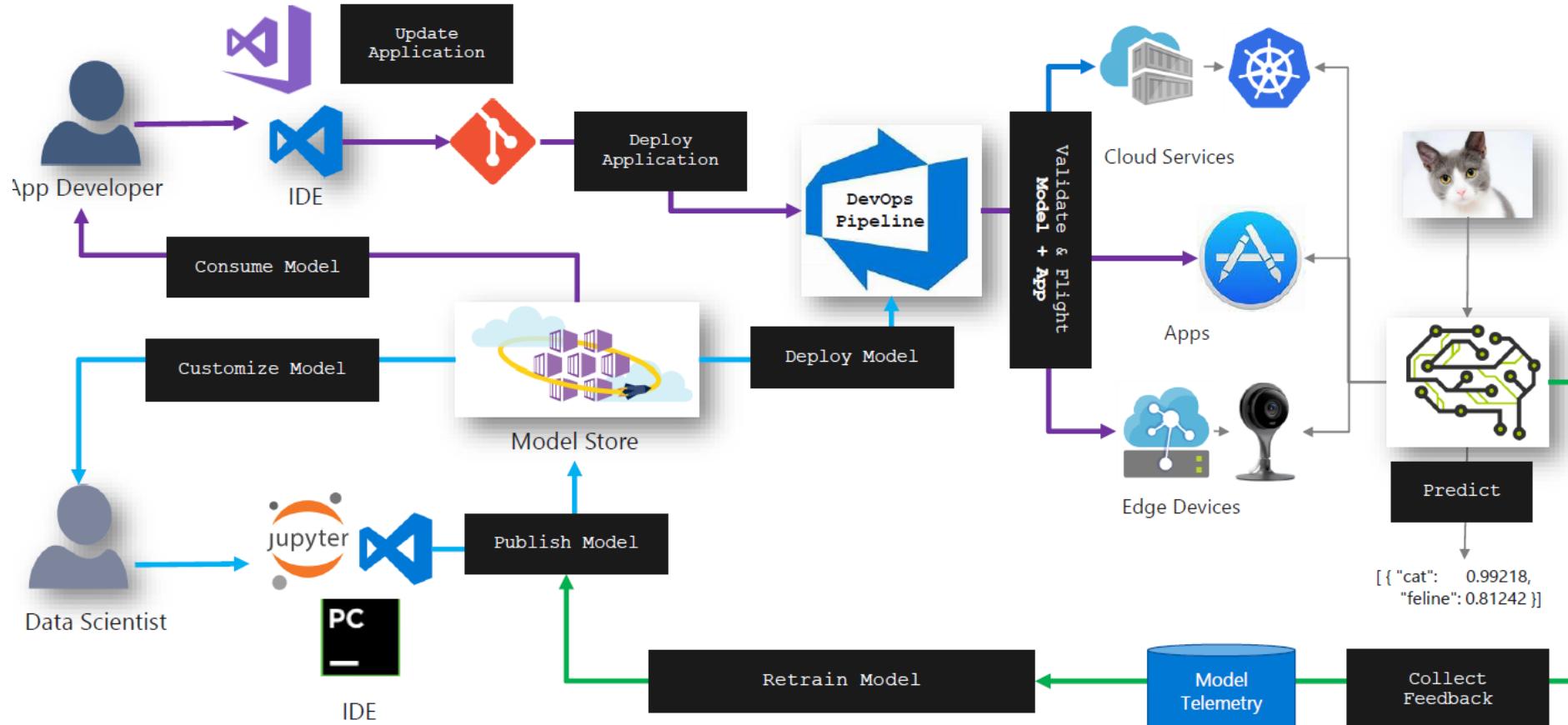


11.1 Integrated AI-Plattforms: Azure

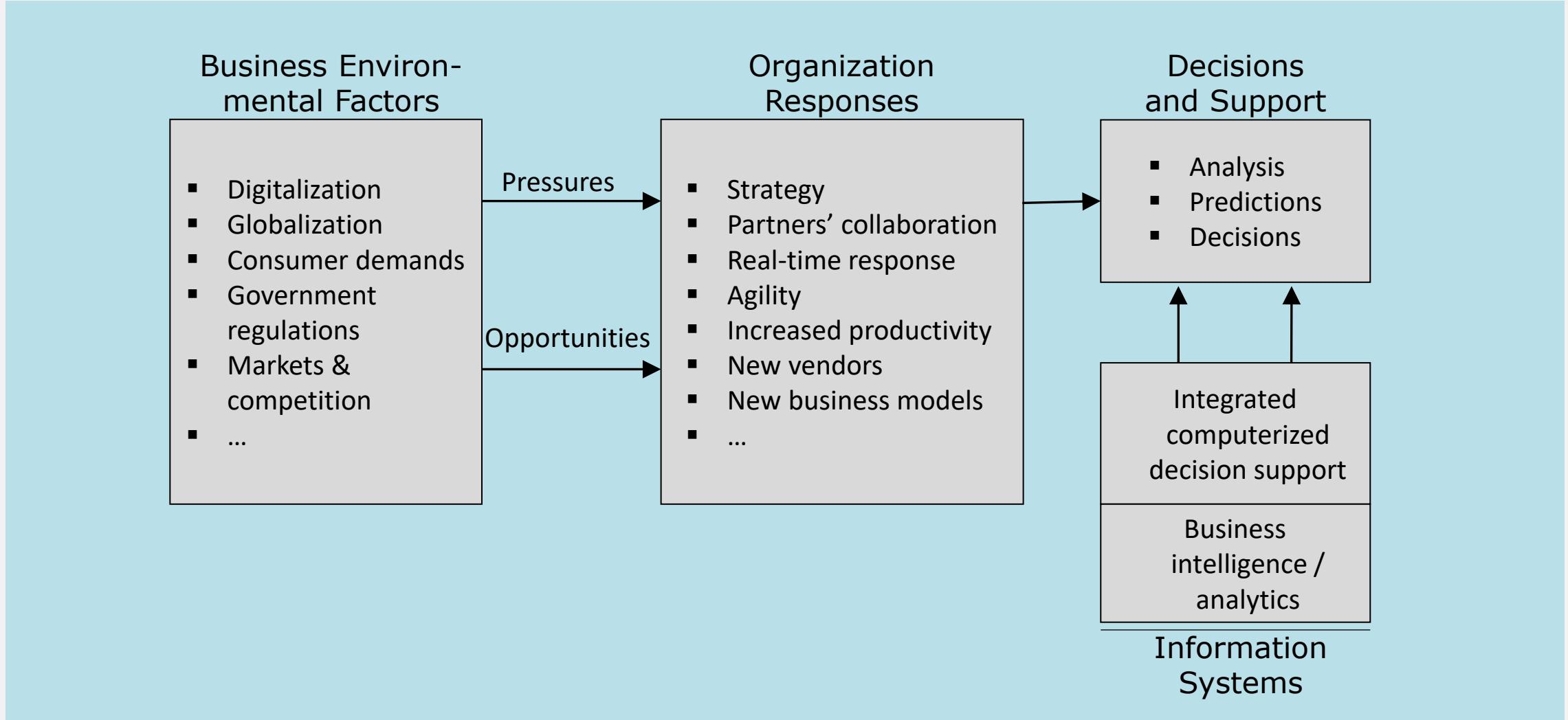


11.1 Example: Azure ML Development Process

ML + App Dev Process



11.1 The Business Pressure-Response-Support Model



Based on Sharda et al. (2014)

11.1 Win by Deployment



Rules of Machine Learning:

Best Practices for ML Engineering

Martin Zinkevich

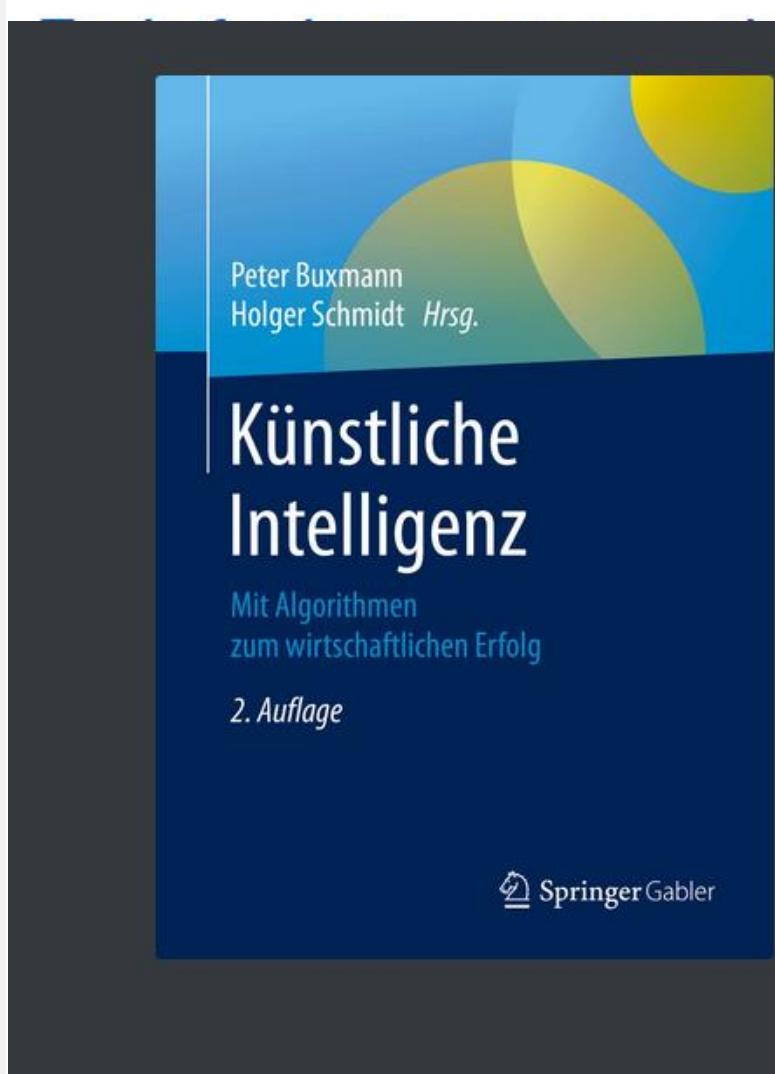
- Deploy simple models in the spirit of winning by shipping
- Make quick iterations through the outermost development loop

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

11.1 Deploy Models on Real Data as Fast as Possible

- 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.

11.1 Successful AI Projects: Book Recommendation



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-Daten durch Machine Learning
- IBM: Intelligente KI-Lösungen
- Empolis: Mit KI immer die richtigen Entscheidungen treffen
- Heraeus: KI-gestützte visuelle Inspektion in der Elektronikindustrie



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



Previous
Exam Task!

Your turn!

Task

Please discuss with your neighbors:

- What is the difference between data understanding and business understanding? Why is there an interaction between these two phases?

Task

Please name three problems, we discussed in lecture, an AI Specialist should have in mind when he/she designs AI-based information systems?

Outline

11

Building Productive AI-based Systems

11.1

The Challenge of Applied AI

11.2

Robotic Agents

11.3

Human-AI Interaction

11.4

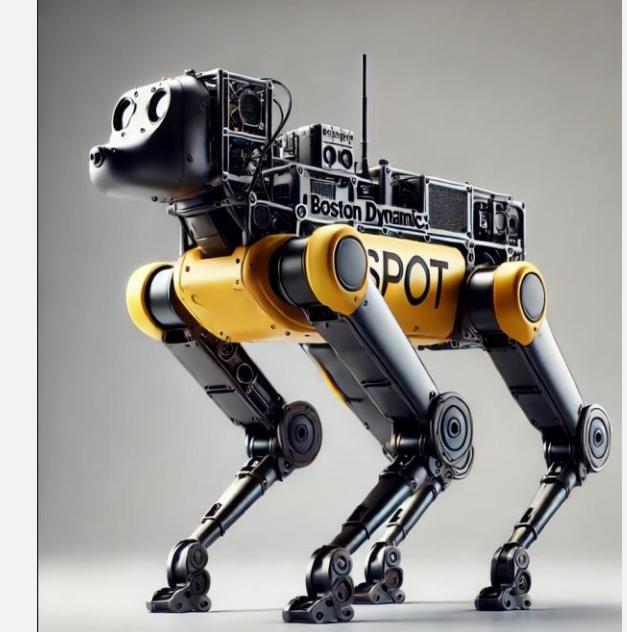
AI Ethics and Governance

11.5

How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge



► 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:

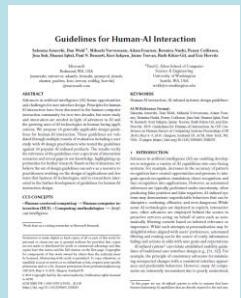
- 135 min + FAQs

► Relevant for Exam:

- 11.1-11.4

11.3 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.



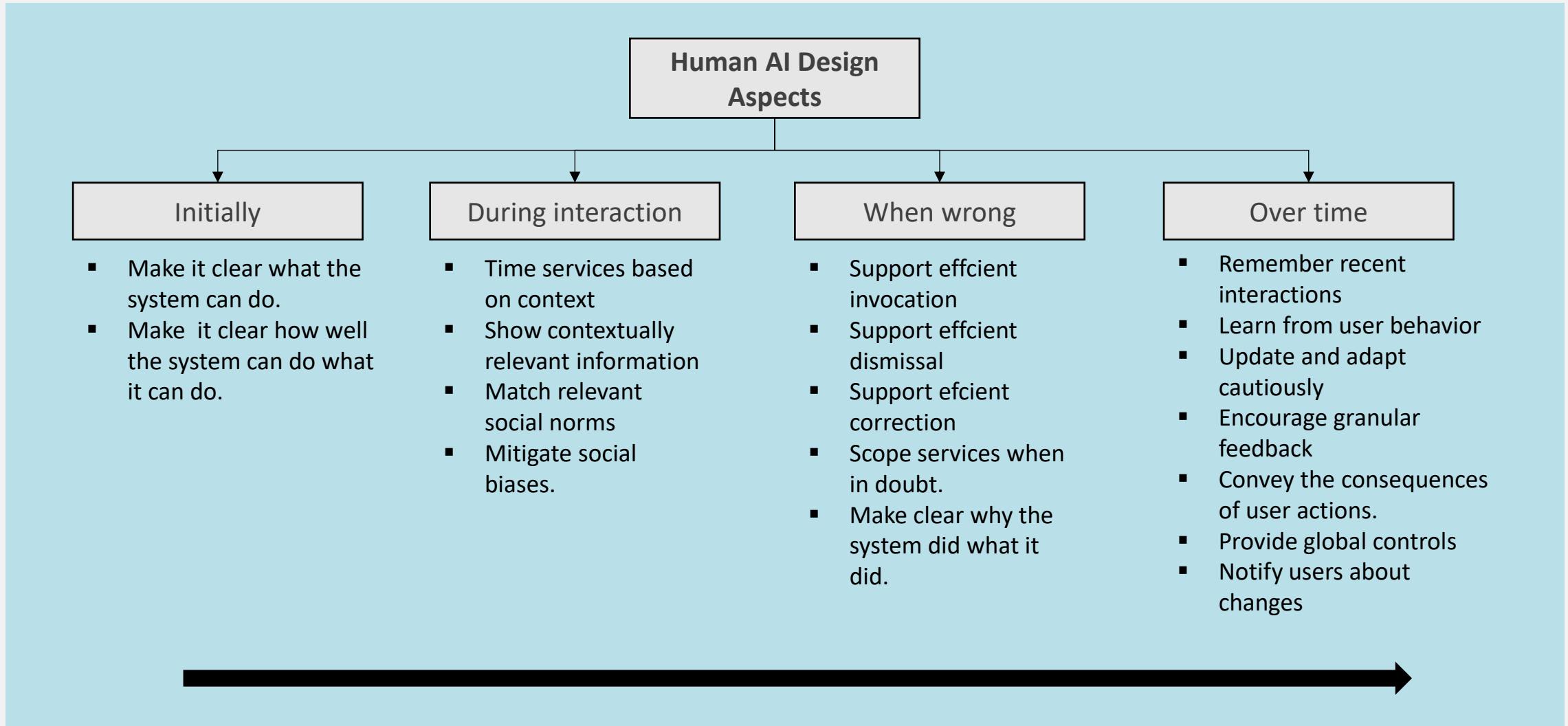
The following part of the lecture is mainly based on Microsoft's Best Practices for AI-Design ↗www.microsoft.com.

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"
When wrong	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."
Over time	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."
	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.

11.3 AI Design Guidelines



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Initially

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)

11.3 AI Design Guideline: Initially

2

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.
- **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)

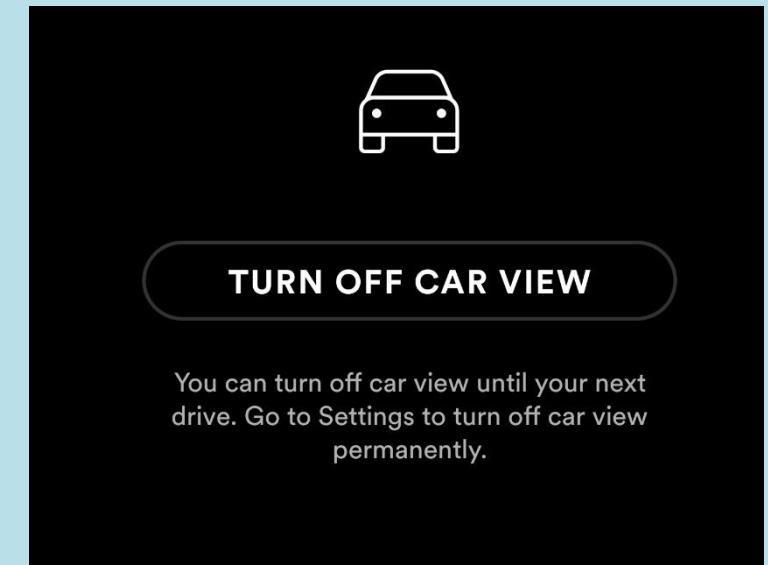
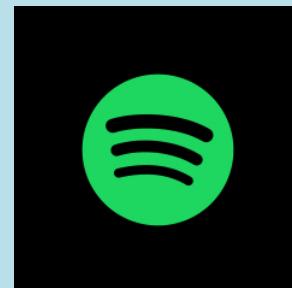
11.3 AI Design Guideline: During interaction

3

G3 - Time services based on context.

- 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



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: During interaction

4

G4 - Show contextually relevant information.

- 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.



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: During interaction

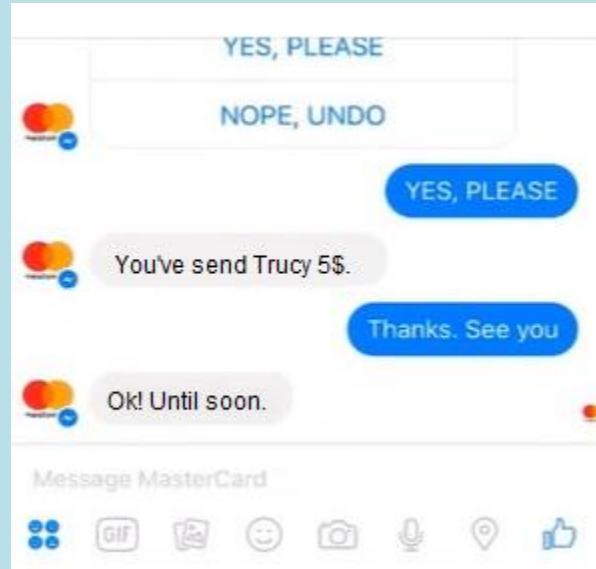
5

G5 - Match relevant social norms

- 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)



Adapted from Amershi, S et al. (2019)

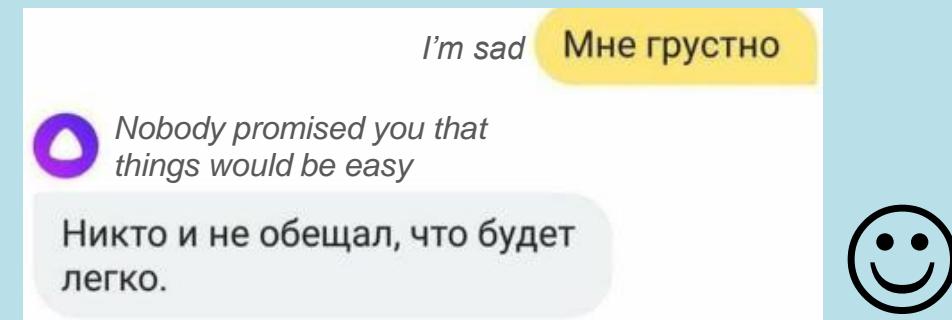
11.3 AI Design Guideline: During interaction

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



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: When wrong

7

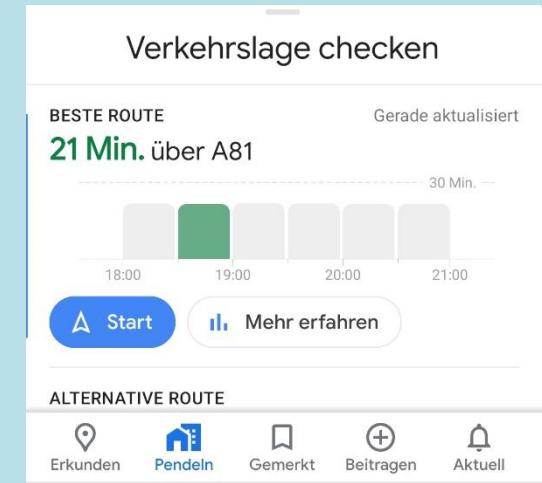
G7 - Support efficient invocation.

- 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)

11.3 AI Design Guideline: When wrong

8

G8 - Support efficient dismissal

- 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."

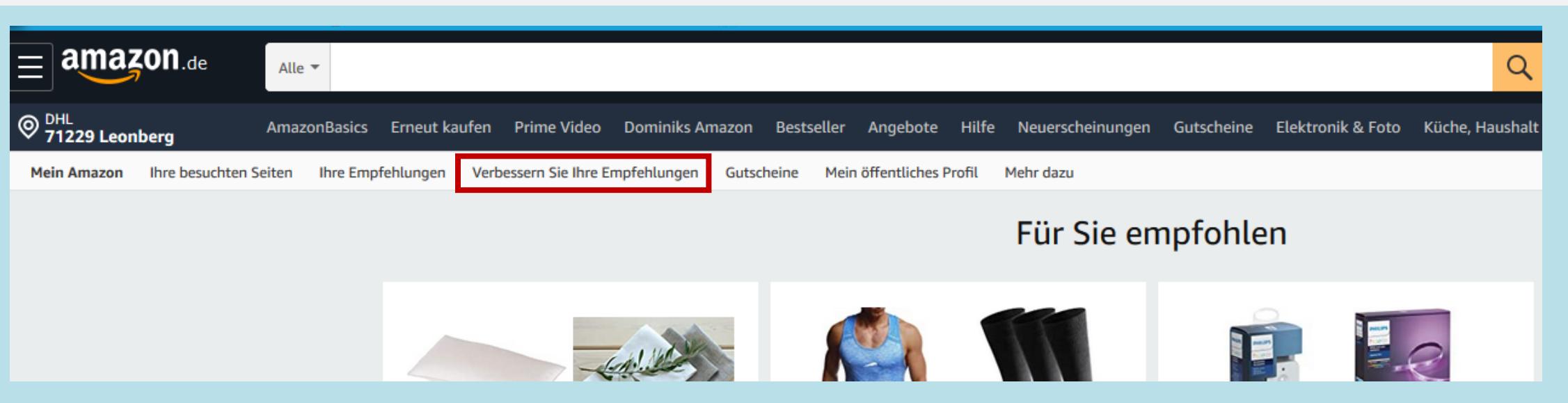
Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: When wrong

9

G9 - Support efficient correction.

- 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)

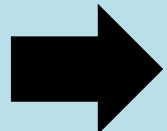
11.3 AI Design Guideline: When wrong

10

G10 - Scope services when in doubt.

- 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)

11.3 AI Design Guideline: When wrong

11

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.
- **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.”

Google News-Hilfe Tippen Sie bitte Ihre Frage ein

Optimieren, was auf Google News personalisiert angezeigt wird

Sie können Google News personalisieren. Dadurch werden Ihnen automatisch mehr Meldungen angezeigt, die für Sie relevant sind, und Sie sehen weniger von denen, die Sie nicht interessieren. Dank der Personalisierung können Sie Inhalte, die Sie interessieren, leichter und schneller sehen.

Indem Sie Google News verwenden, teilen Sie dem Dienst mit, was Sie lesen möchten. Je mehr Sie den Dienst verwenden, desto besser wird der Bereich "Für mich". Sie können Google News auch direkt mitteilen, was Ihnen gefällt.

Hinweis: Diese Funktion ist nur verfügbar, wenn Sie in Ihrem Google-Konto angemeldet sind.

Für mich
Empfohlen auf Basis deiner Interessen

Zins-Attacke: DKB liebäugelt mit revolvierenden Kreditkarten
finanz-szene.de · Gestern

Mehr zum Thema "Für mich"

Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

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."



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

13

G13 - Learn from user behavior.

- 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."

The screenshot shows a user's Amazon Prime account dashboard. At the top, there's a navigation bar with a search icon, a German flag, a greeting 'Hallo, Dominik', a 'Konto und Listen' dropdown, links for 'Warenrücksendungen und Bestellungen', 'Entdecken Sie Prime', and a shopping cart icon with '0' items labeled 'Einkaufswagen'. Below the navigation, there are four main sections: 'Nochmals kaufen' (highlighted with a red box), 'Meine Listen', 'Mein Konto', and a promotional banner for 'Amazon Prime | 30 Tage gratis nutzen'. The 'Nochmals kaufen' section displays a product thumbnail for 'PARADIES Kopfkissen Softy-Tip medium 40 x 80 cm' priced at '28,00 €' with a 'In den Einkaufswagen' button. The 'Meine Listen' section lists 'Alexa-Einkaufsliste' (0 Artikel), 'Wunschzettel', 'Einkaufsliste', 'Bücher und Hörbücher', and 'Ihre Listen'. The 'Mein Konto' section lists 'Mein Konto', 'Meine Bestellungen', 'Meine Dash Buttons', 'Wunschzettel', 'Empfehlungen', and 'Spar-Abo verwalten'.

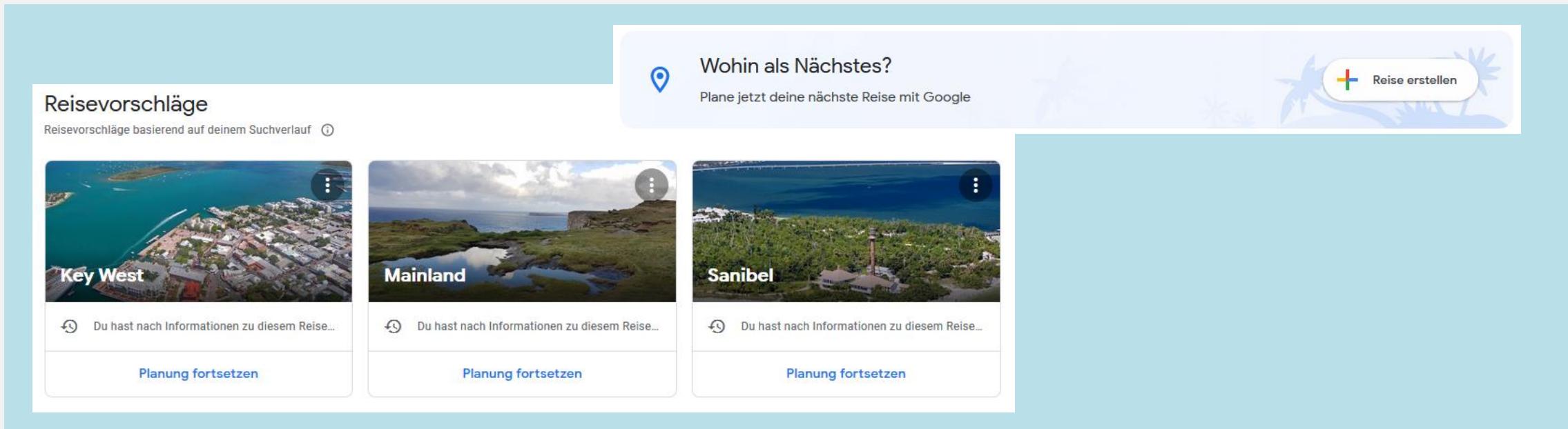
Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

14

G14 - Update and adapt cautiously

- 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)

11.3 AI Design Guideline: Over time

15

G15 - Encourage granular feedback.

- 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.”



Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

16

G16 - Convey the consequences of user actions.

- 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.”



Google

decision support sys|

X

- decision support system
- decision support systems
- decision support system example
- decision support system beispiel
- decision support system deutsch

Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

17

G17 - Provide global controls

- 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.”

Wohin als Nächstes?

Plane jetzt deine nächste Reise mit Google

Nur für dich sichtbar ⓘ

Orte

Pfullingen Kilchberg Feldberg (Schwarzwald)

Rust Santa Clara Ettlingen

Wien Hué Nh  Trang

Adapted from Amershi, S et al. (2019)

11.3 AI Design Guideline: Over time

18

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.”



Adapted from Amershi, S et al. (2019)

11.3 Classroom Task



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

11

Building Productive AI-based Systems

11.1

The Challenge of Applied AI

11.2

Robotic Agents

11.3

Human-AI Interaction

11.4

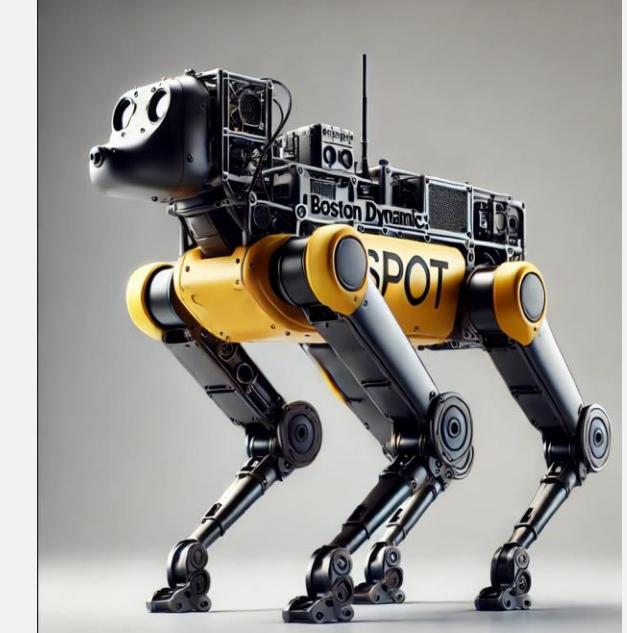
AI Ethics and Governance

11.5

How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge



► 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:

- 135 min + FAQs

► Relevant for Exam:

- 11.1-11.4

11.4 Amazon AI Recuriter doesn't like Women

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.

11.4 Biased AI Models Have Real-World Consequences

YT Fact Check Algorithm Fails

Mail Online

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

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



Google Photo Fail 2015

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

A screenshot of a Twitter post from June 29, 2015. The post reads "Google Photos, y'all fucked up. My friend's not a gorilla." and includes a link. Below the text are three small images: one of a person on a bike labeled "Bikes", one of two people labeled "Gorillas", and one of a person in a graduation gown labeled "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

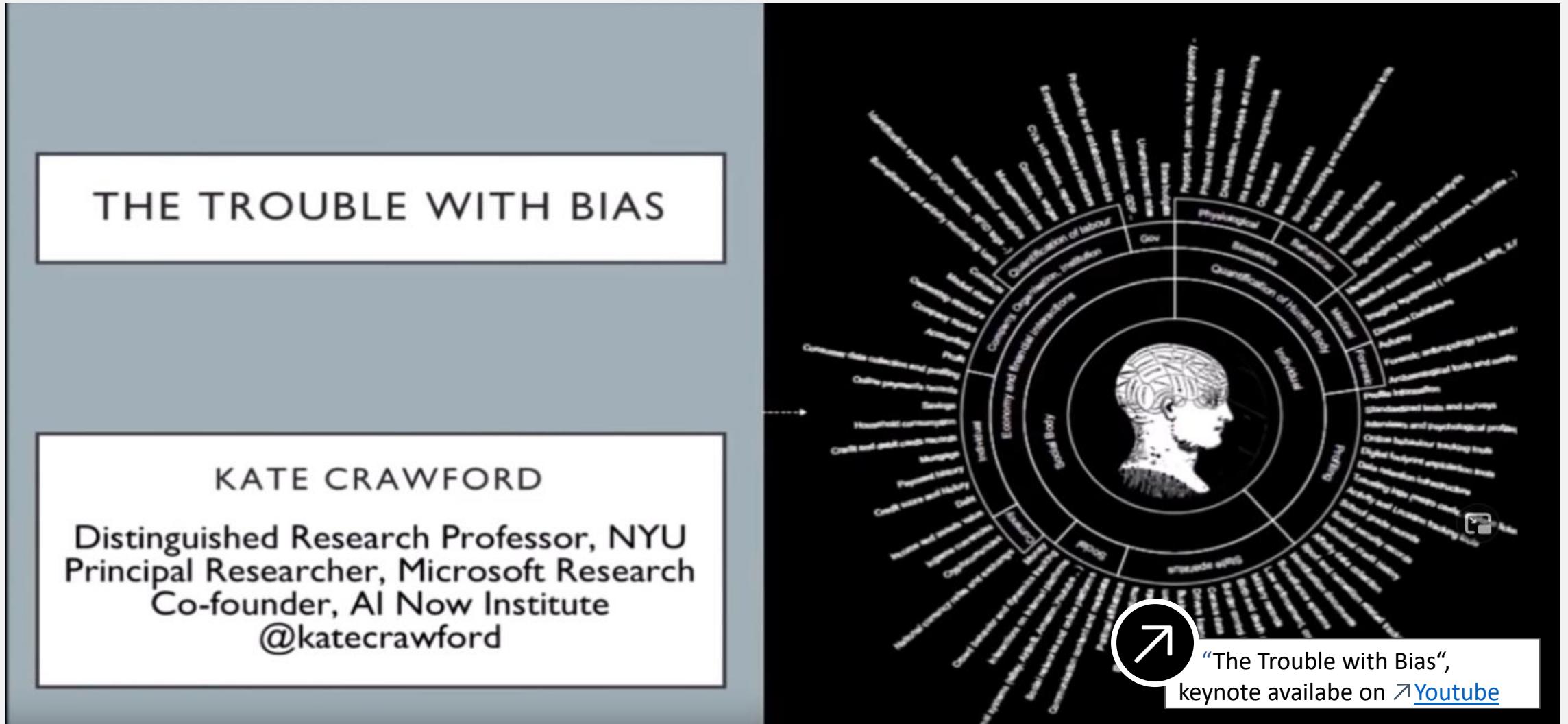
Image source: COMPAS ↗[Pixabay](#) (2019) / ↗[CC0](#)

11.4 Biased Data is a Huge Challenge in AI Design



Image source: [Associate Professor Kate Crawford, University of New South Wales](#) (2009) by [andresmh](#) from Wikimedia / [CC BY-SA 2.0](#)

11.4 The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford



11.4 Harm that AI Can Cause

- 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

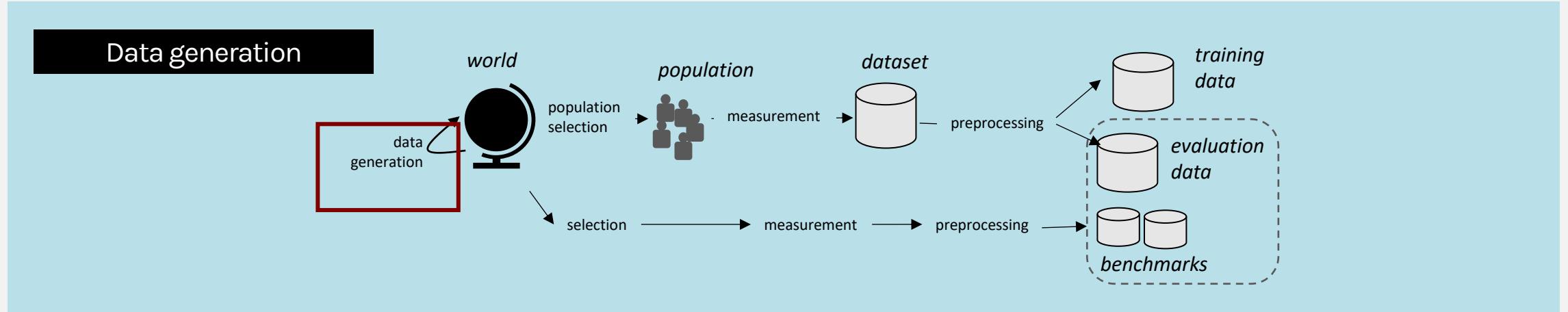
11.4 Ethics in Artificial Intelligence

- **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

11.4 Ethical Issues in Artificial Intelligence

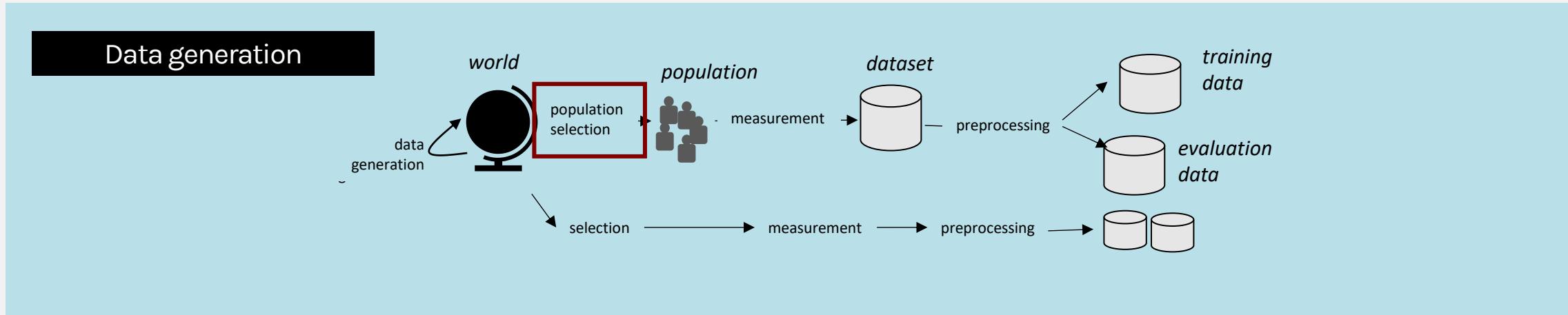
- **What AI is**
 - Bias and fairness
 - Accountability
 - Transparency and understandability
 - Interpretability
 - Decision-maker/support
- **What AI does**
 - Work- and process security
 - Cyber-security and malicious use
 - Privacy
 - Human-AI interaction
- **What AI impacts**
 - Jobs and labor trends
 - Impact to modern society
 - Human-human interaction
 - Communication
 - Law
 - Politics
- **What AI can be**
 - Singularity?
 - Robot rights (see e.g Assimov)?
 - Part of society?

11.4 Historical Bias



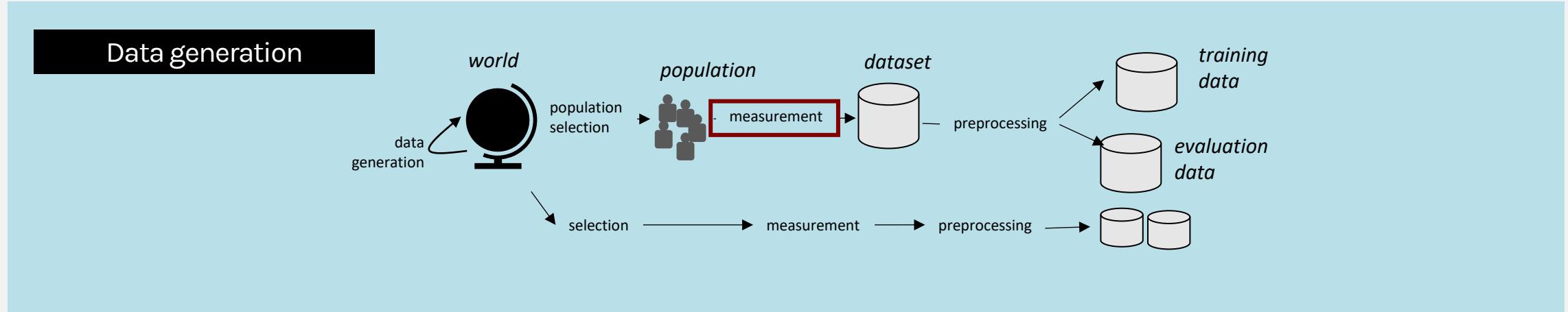
- Historical bias arises when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model.
- It is a normative concern with the state of the world, and exists even given perfect sampling and feature selection.

11.4 Representation Bias



- 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

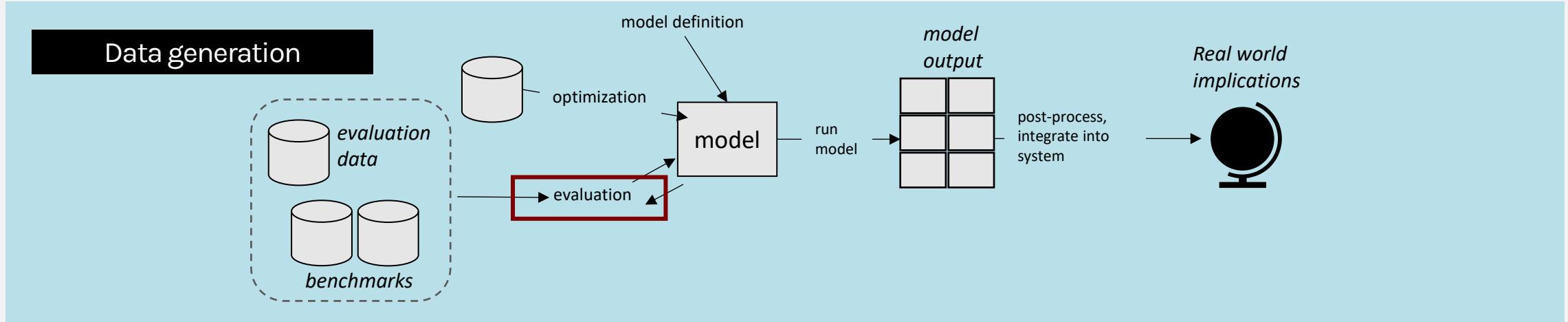
11.4 Measurement Bias



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.

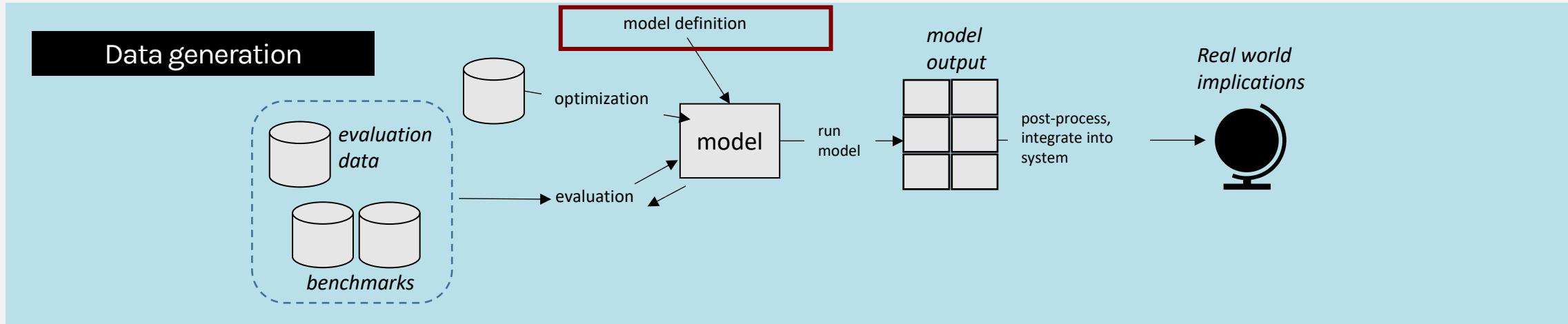
11.4 Evaluation Bias



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

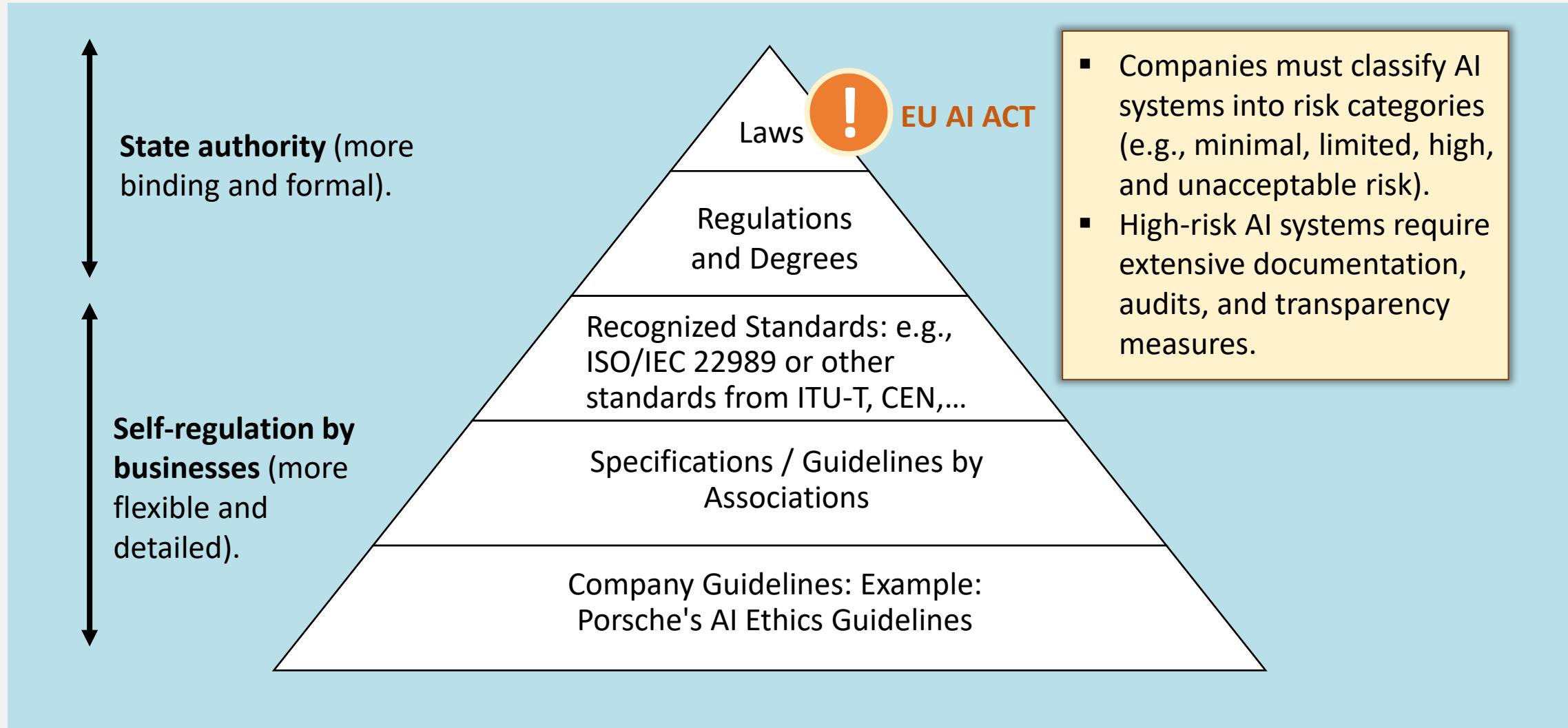
- Evaluation bias can also arise from the use of performance metrics that are not granular or comprehensive enough.

11.4 Aggregation Bias



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.

11.4 Regulation of AI at Different Levels in the European Union



Adapted from the Artificial Intelligence Act - Regulation (EU) 2024/1689.

11.4 EU AI Act Risk Categories

EU AI Act Risk Categories

Risk tiers	Prohibited	High-risk	Low-risk
	<p>Obligations: Banned, although narrow exceptions may be allowed (e.g., for law enforcement)</p> <p>Examples: AI systems circumventing users' free will or used for "social scoring"</p>	<p>Obligations: Compliance with strict requirements</p> <p>Examples: Most deployments of AI technology with impact on people will fall in this category</p>	<p>Obligations: Lighter requirements, such as inventory and (other) documentation</p> <p>Examples: AI-enabled recommender systems, spam and virus filters</p>
Requirements	Risk assessments, transparency obligations and registration in an EU-wide database will apply to specific AI-enabled applications.		
	Additional mandatory requirements will apply to the highest-performance General Purpose AIs that entail systemic risks.		

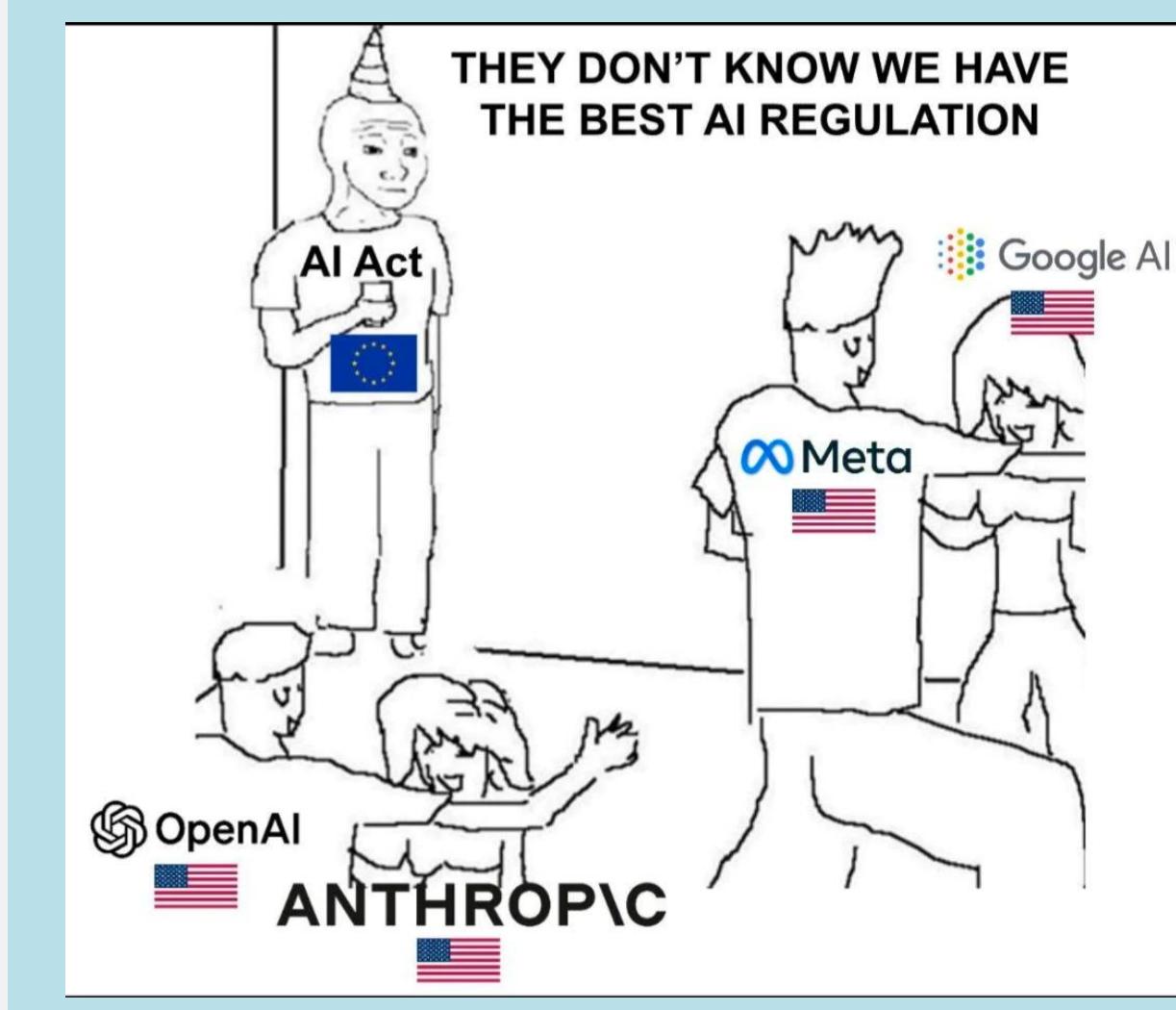
Adapted from the Artificial Intelligence Act - Regulation (EU) 2024/1689.

11.4 Implications for AI Development in the EU

- Development costs rise due to mandatory compliance processes like data quality checks, risk assessments, and third-party evaluations
- The EU's stringent rules may place its companies at a disadvantage compared to less-regulated regions (e.g., US, China)
- Companies must maintain extensive records of training data, model design, and decision-making processes
- High fines and penalties: Non-compliance can lead to significant financial penalties (up to €30M or 6% of global annual turnover)!

Adapted from the Artificial Intelligence Act - Regulation (EU) 2024/1689.

11.4 The EU AI Act: Pioneering Leadership or Strategic Misstep?



The Guardian

News | Opinion | Sport | Culture | Lifestyle | Eur

World Europe US Americas Asia Australia Middle East Africa Inequality Global development

Apple delays launch of AI-powered features in Europe, blaming EU rules

THE TIMES

Meta presses pause on launch of latest AI in Britain and EU

The Facebook owner has blamed the stalled rollout of Meta AI in Europe on uncertainty around what data can be used to train AI models.



Your turn!

Task

Consider the following aspects, what are ethical problems and possible biases in AI usage and AI design.

- Automation of communication (e.g. Case Ashley Madison)
- Homework automation/solutions with ChatGPT
- Generation of fake image and videos (e.g. deep fakes)



Image source: [↗ Pixabay](#) (2019) / [↗ CCO](#)

Outline

11

Building Productive AI-based Systems

11.1

The Challenge of Applied AI

11.2

Robotic Agents

11.3

Human-AI Interaction

11.4

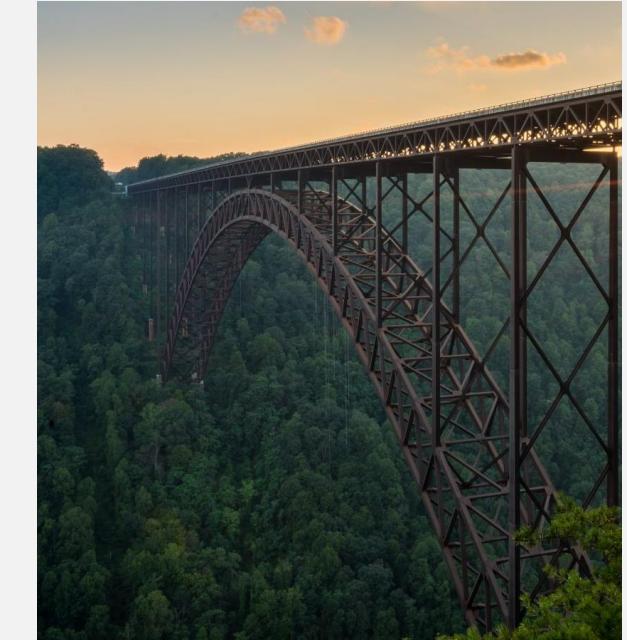
AI Ethics and Governance

11.5

How to Continue your AI Journey

Exam Preparation and Course Repetitorium

Capstone Project / Case Challenge



► 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:

- 135 min + FAQs

► Relevant for Exam:

- 11.1-11.4

11.5 Start your AI Journey



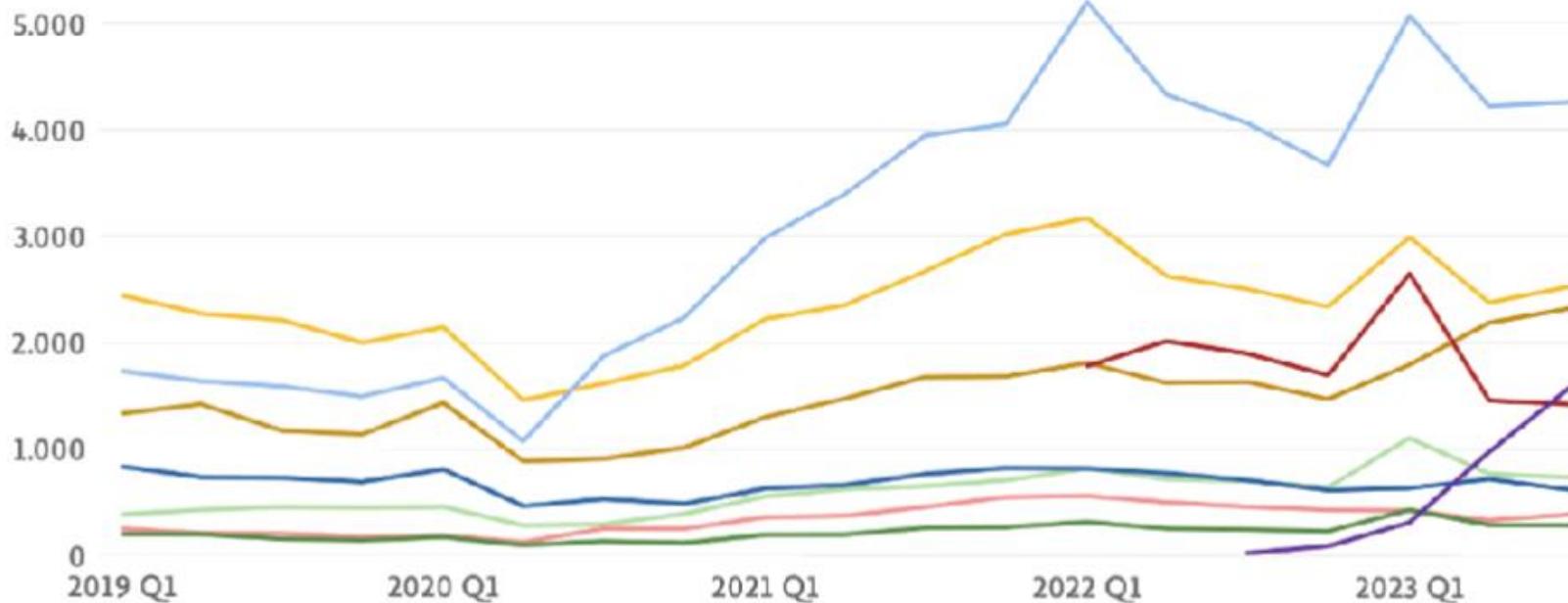
Image source: ↗ [Pixabay](#) (2019) / ↗ [CC0](#)

11.5 How is the Job Situation?

Stellenausschreibungen für KI-Experten

Ausgeschriebene Positionen in Deutschland für folgende Fachgebiete

Künstliche Intelligenz Machine Learning Deep Learning Data Scientists / Analysts Data Engineer Computer Vision Virtuelle Agenten Robotic Process Automation Generative KI



Adapted from FAZ PRO, D:Economy (2023), based on saco / Indec Gruppe 2023

Wirtschaft

FORSCHUNGSGIPFEL IN BERLIN

„Der Arbeitsmarkt für KI-Experten ist leergefegt“

von DIETMAR HARHOFF

VERÖFFENTLICHT AM 19.03.2019 - 10:18



Google und die anderen Tech-Giganten zahlen KI-Profis hohe Gehälter. Die öffentliche Forschung sollte ihnen nicht das Feld überlassen, schreibt der Vorsitzende der Expertenkommission Forschung und Inn in einem Gastbeitrag.



11.5 Untold Truths about AI Jobs (personal opinion !)

- Big gap between theoretical vs. practical work (aka university vs. industry)
- There are many misleading learning resources
- A solid background in the foundations (e.g. linear algebra) is crucial
- Productization is Key
- Neural Networks are not the silver bullet
- Challenges Often Aren't Technical
- Impact Over Skills List



TABULAR DATA: DEEP LEARNING IS NOT ALL YOU NEED

Ravid Shwartz-Ziv
ravid.ziv@intel.com
IT AI Group, Intel

Amitai Armon
amitai.armon@intel.com
IT AI Group, Intel

November 24, 2021

ABSTRACT

A key element in solving real-life data science problems is selecting the types of models to use. Tree ensemble models (such as XGBoost) are usually recommended for classification and regression problems with tabular data. However, several deep learning models for tabular data have recently been proposed, claiming to outperform XGBoost for some use cases. This paper explores whether these deep models should be a recommended option for tabular data by rigorously comparing the new deep models to XGBoost on various datasets. In addition to systematically comparing their performance, we consider the tuning and computation they require. Our study shows that XGBoost outperforms these deep models across the datasets, including the datasets used in the papers that proposed the deep models. We also demonstrate that XGBoost requires much less tuning. On the positive side, we show that an ensemble of deep models and XGBoost performs better on these datasets than XGBoost alone.

11.5 Recap - Job Profiles in AI (Examples)

AI Specialist

- Professional responsible for AI related topics at companies
- Designing, developing and maintaining simple AI related solutions
- Product owner in AI-related projects

Data Engineer

- Models scalable database and data flow architectures
- Develops and improves the IT infrastructure on the hardware and software side
- Deals with topics such as IT Security , Data Security and Data Protection

Data Scientist

- AI-Expert responsible for collecting, analyzing and interpreting extremely large amounts of data
- The role is an offshoot of several traditional technical roles, including mathematician, scientist, statistician and computer professional

Robotics Scientist

- Engineer responsible for implementing intelligent robots
- Bridge between mechanical engineering, electrical engineering, computer science

Business Data Analyst / BI Developer

- Designing, developing and maintaining business intelligence solutions
- Crafting and executing queries upon request for data
- Presenting information through reports and visualization

Machine Learning Engineer

- Engineer responsible for implementing intelligent robots
- Bridge between mechanical engineering, electrical engineering, computer science

AI Research Scientist

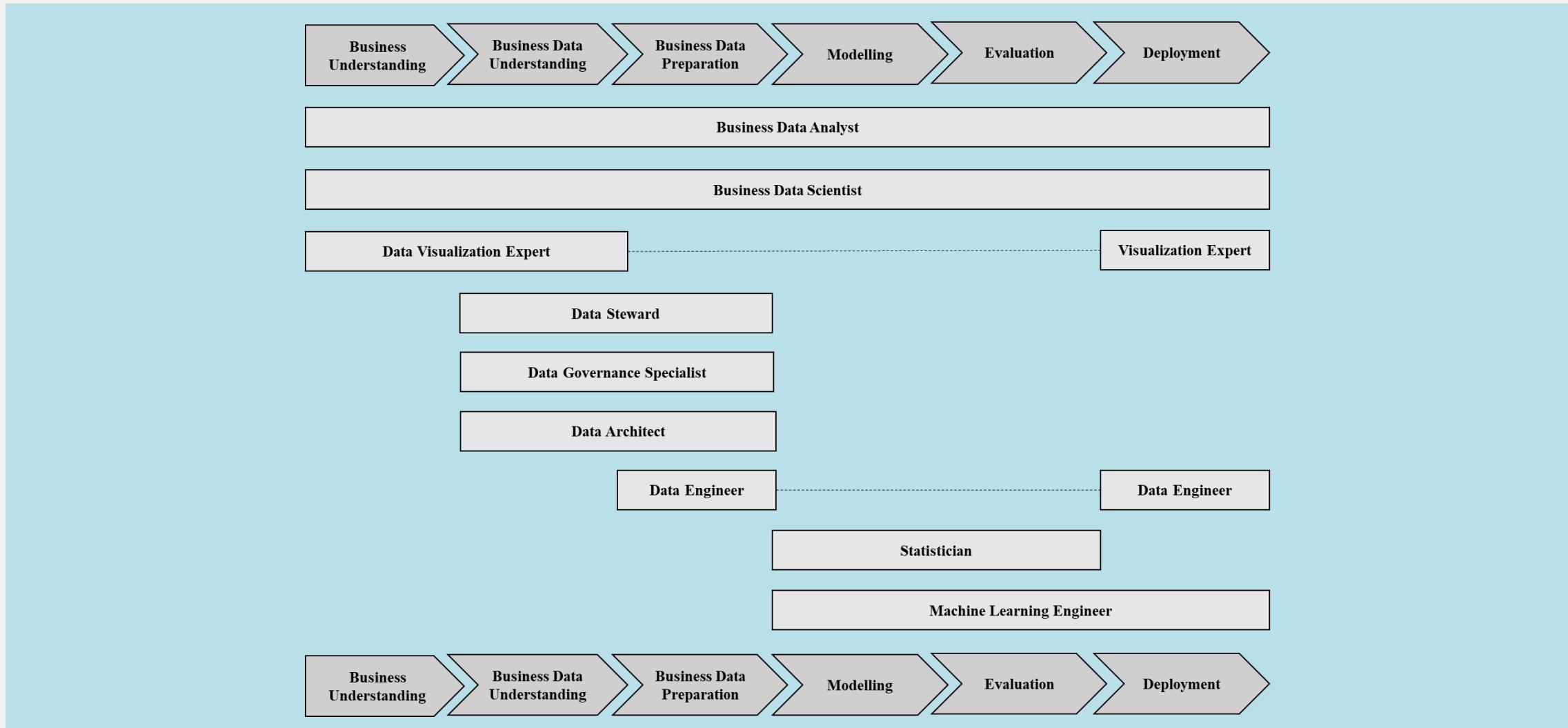
- Works mostly for universities or big companies
- Passionately drive and further advance innovations in the field of computer vision

AI Product Owner

- Defining the vision and strategy of AI-driven products
- Ensuring that the development team creates solutions that meet business goals and user needs.

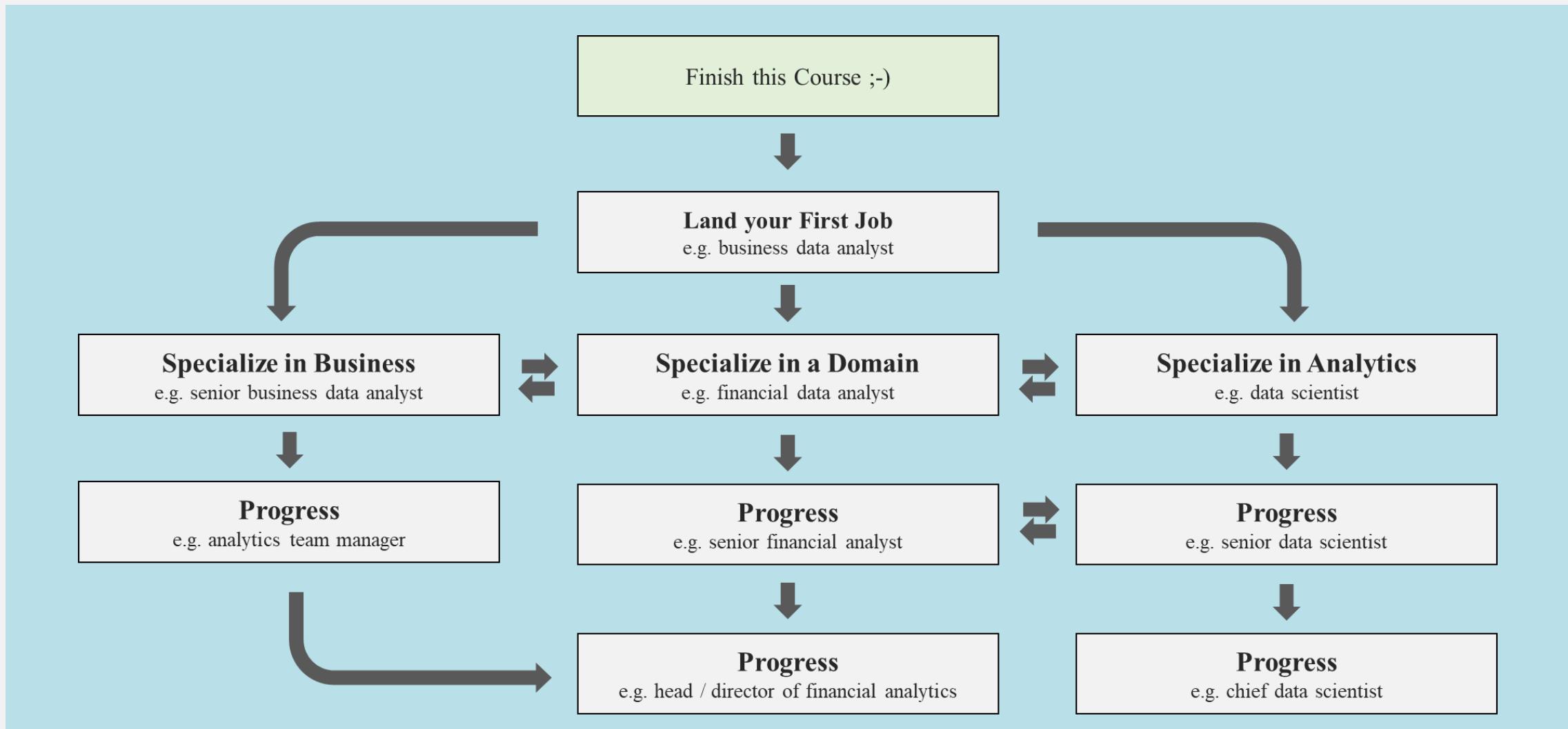
...

11.5 Possible Job Roles Across an AI Project



Adapted from Jung, D. (2024)

11.5 Possible Career Paths in the Industry



Adapted from Jung, D. (2024)

11.5 Getting the Right Experience – University Courses

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

Computer Science

- Data Management and Warehousing
- Software Engineering and Architecture
- Algorithms

Statistics and Math

- Econometrics
- Operations Research

Specialization in AI

- Natural Language Processing,
- Robotics,
- Machine Learning (!)
- ...

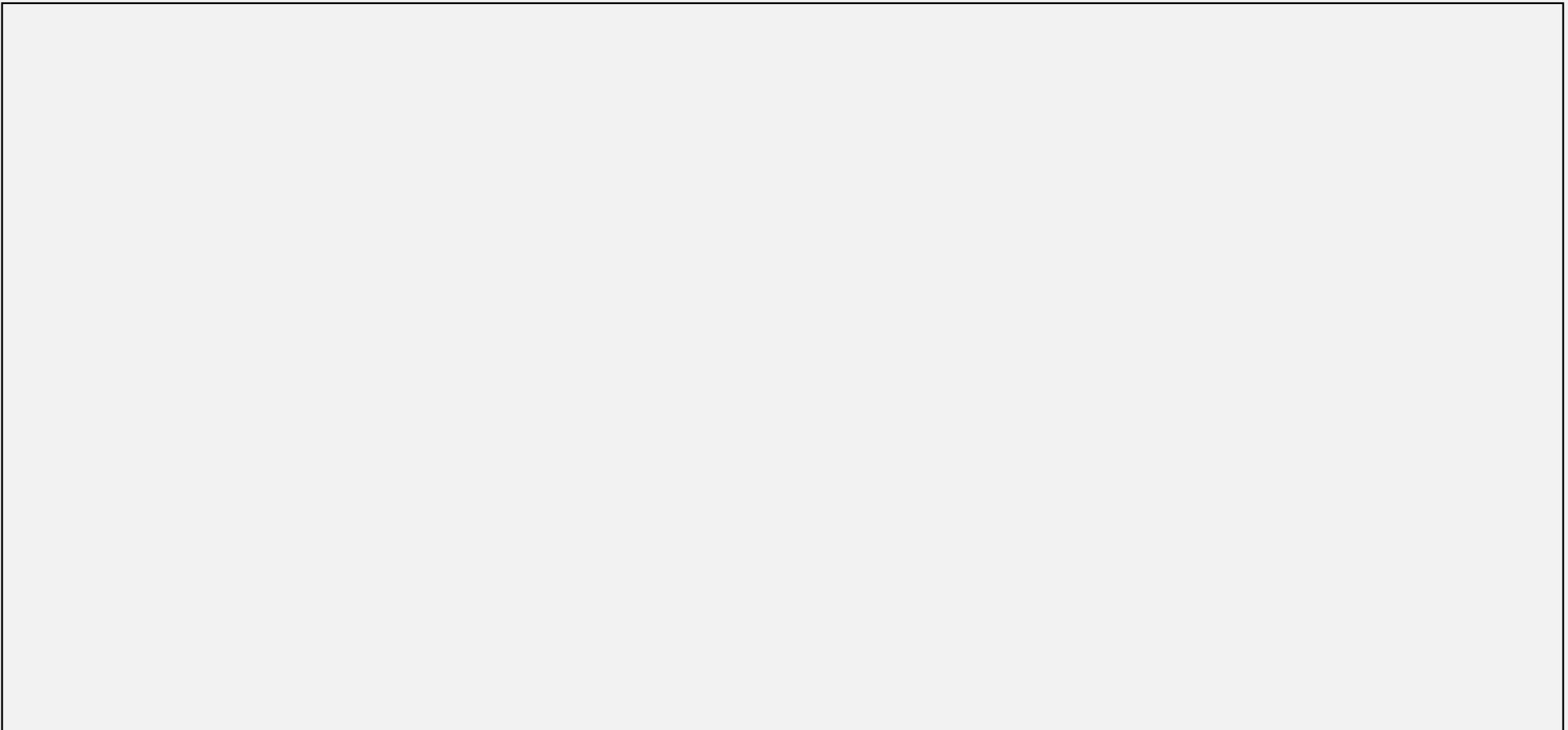
Or Combine with other domains:

- Bio-Informatics
- Finance
- ...



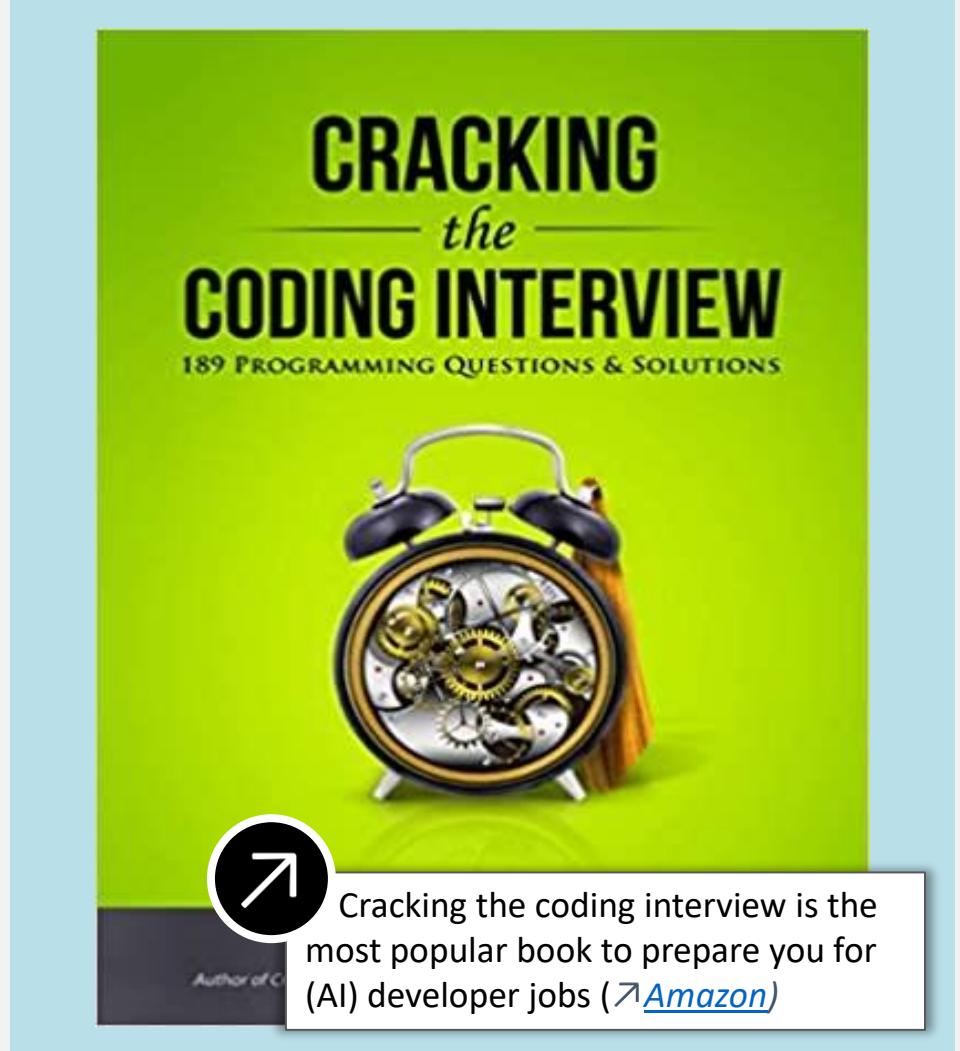
You have questions related to your curriculum? Just feel free to ask now or come to me after lecture!

11.5 Getting the Right Experience – PhD or Not



11.5 Preparing the Technical Interview

- 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



11.5 Maintaining Your Competitive Advantage is Hard

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

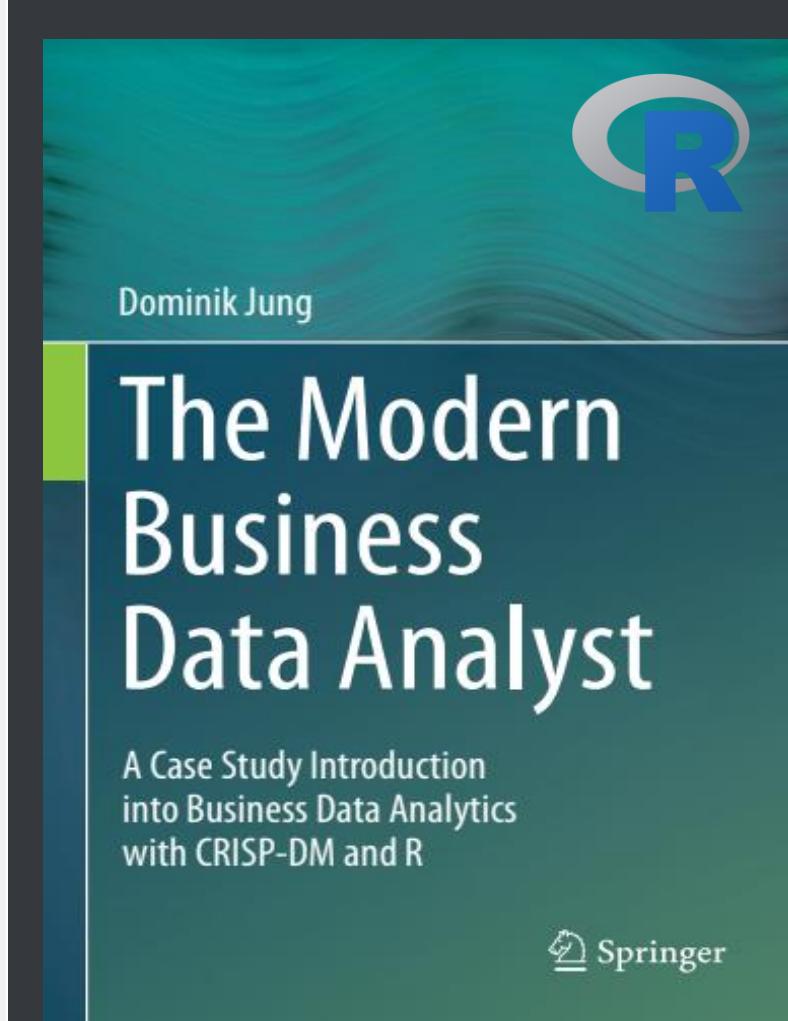
Make pie charts as a CEO: \$14 million/year

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 solely on your AI Skills

11.5 Book Recommendation if You Want to Become Business Data Analyst



Contents:

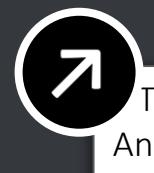
- Starting-point for data science beginners and data analysts
- Hands-on examples and real business problems with R
- Accompanying course website and business cases to practice

Accompanying course:



Business Data Analytics and Decision Support with R

↗ course materials



The Modern Business Data Analyst (↗[Amazon](#) | ↗[Springer](#))

Your next project: AI Capstone Challenge

The collage includes:

- A top-left image showing two people working at a desk, one wearing glasses.
- A top-right image showing a person's hands working on a laptop keyboard.
- A middle section featuring a diagram titled "Aftersales.digital ecosystem" with a red border. It includes a car icon and the text "SEAMLESS JOURNEY".
- A bottom section showing a grid of 12 video call thumbnails of people, likely participants in a virtual meeting.

*Example: 2021 Capstone
with Warranty and
Goodwill teams
@Porsche AG*

11. Exercises

Workbook Exercises

- Please read the chapter 25 from Russell, S., & Norvig, P. (2016) to understand how to build and deploy agents and robotic AI-based systems. Then continue with chapters 26 to 27 and reflect the pros and cons of AI for i) your life, ii) everyday's life iii) application in industry iv) civilization. Then work through the exercises of the chapters to prepare for the exam.

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 ideas, wishes, or questions please feel free to write me.



You will do well in your future
endeavors!



Microsoft AI ft. Common

<https://www.youtube.com/watch?v=9tucY7Jhhs4>

Artificial Intelligence: Algorithms and Applications with Python - Dr. T



11. 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. Feng, S., Park, C. Y., Liu, Y., & Tsvetkov, Y. (2023). From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models.
4. Google (2019a): Google's People+AI Guidebook - Designing human-centered AI products. Online available at: <https://pair.withgoogle.com/guidebook>
5. Jung, D. (2024) The Modern Business Data Analyst: A Case Study Introduction into Business Data Analytics with CRISP-DM and R
6. McDowell, G. L. (2015). Cracking the Coding Interview: 189 programming questions and solutions.
7. Orosz, G. (2024). The Software Engineer's Guidebook: Navigating senior, tech lead, and staff engineer positions at tech companies and startups. Pragmatic Engineer BV.
8. 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>
9. The Artificial Intelligence Act - Regulation (EU) 2024/1689. Online available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>
10. 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>

11. References

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-scaps-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>

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 interesting tutorials and best practices you find on Google Developers 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
- For an overview of political biases in LLM see e.g. Feng, S et al (2023).

11. Google's People+AI Guidebook (Google Research, Google Design)

+PAIR

GUIDEBOOK EXPLORABLES TOOLS RESEARCH



People + AI Guidebook

The People + AI Guidebook was written to help user experience (UX) professionals and product managers follow a human-centered approach to AI.

Getting Started

Its recommendations are based on data and insights from over a hundred individuals across Google product teams, industry experts, and academic research.

These six chapters follow the product development flow, and each one has a related worksheet to help turn guidance into action.

User Needs + Defining Success

Identify user needs, find AI opportunities, and design your reward function.

[Read Chapter](#) [Get Worksheet](#)

Data Collection + Evaluation

Decide what data are required to meet your user needs, source data, and tune your model.

[Read Chapter](#)

Mental Models

Explainability + Trust

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

11. Glossary

AI Bias	Systematic errors in AI systems that result in unfair or prejudiced outcomes due to biases in training data or algorithm design
CRISP-DM	A widely used framework for developing data science projects, comprising stages like business understanding, data understanding, data preparation, modeling, evaluation, and deployment.
Explainable AI	<i>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</i>
EU AI Act	A regulatory framework by the European Union aimed at categorizing AI systems based on risk levels and imposing compliance measures, with significant implications for businesses operating in the EU.
Lifecycle Approach	A structured methodology for AI project development, spanning the inception of an AI use case through to deployment and maintenance.
XAI / Human-AI Interaction	Guidelines and practices for designing AI systems that effectively interact with users, covering aspects like user trust, error handling, and adaptability over time.

Bucket List - US National Parks

1872 Yellowstone	1944 Big Bend	2000 Cuyahoga Valley (National Recreation Area 1974)
1890 Yosemite (CA SP 1864) Sequoia	1956 Virgin Islands	2003 Congaree (Congaree Swamp NM 1976)
1899 Mount Rainier	1962 Petrified Forest (NM 1906)	2004 Great Sand Dunes (NM 1932)
1902 Crater Lake	1964 Canyonlands	2013 Pinnacles (NM 1908)
1903 Wind Cave	1966 Guadalupe Mountains (established 1972)	2018 Gateway Arch (Jefferson National Expansion Memorial 1935)
1906 Mesa Verde	1968 North Cascades Redwood	2019 Indiana Dunes (National Lakeshore 1966) White Sands (NM 1933)
1910 Glacier	1971 Capitol Reef (NM 1937) Voyageurs (established 1975)	2020 New River Gorge (National River 1978)
1915 Rocky Mountain	Arches (NM 1929)	
1916 Lassen Volcanic (Cinder Cone NM & Lassen Peak NM 1907) Hawaii Volcanoes (originally part of Hawaii NP)	1978 Theodore Roosevelt (National Memorial Park 1947) Badlands (NM 1929)	
Haleakala (originally part of Hawaii NP)	1980 Channel Islands (NM 1938) Biscayne (NM 1968)	
1917 Denali (originally Mt McKinley NP, renamed in 1980)	Katmai (NM 1918)	
1919 Grand Canyon (NM 1908) Zion (Mukuntuweap NM 1909) Acadia (Sieur de Monts NM 1916, originally Lafayette NP, renamed in 1929)	Glacier Bay (NM 1925)	
1921 Hot Springs (Reservation 1832)	Gates of the Arctic (NM 1978)	
1926 Shenandoah (established 1935)	Kenai Fjords (NM 1978)	
1928 Bryce Canyon (NM 1923)	Kobuk Valley (NM 1978)	
1929 Grand Teton	Lake Clark (NM 1978)	
1930 Carlsbad Caverns (1923)	Wrangell St. Elias (NM 1978)	
1934 Everglades (established 1947) Great Smoky Mountains (established 1940)	1986 Great Basin (Lehman Caves NM 1922)	
1938 Olympic (Mt Olympus NM 1909)	1988 National Park of American Samoa (established 1993)	
1940 Kings Canyon Isle Royale	1992 Dry Tortugas (Fort Jefferson NM 1935)	
1941 Mammoth Cave	1994 Death Valley (NM 1933)	
	Saguaro (NM 1933)	
	Joshua Tree (NM 1936)	
	1999 Black Canyon of the Gunnison (NM 1933)	