

WIE KOMMEN DIE DATEN IN DIE KI?

Theoretische Einführung und Best Practices für die Nutzung von KI auf Basis von Unternehmensdaten

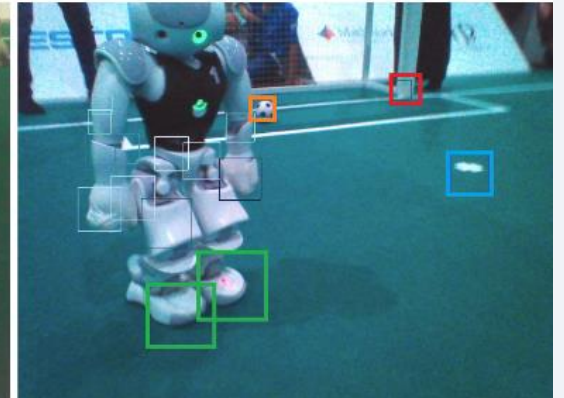
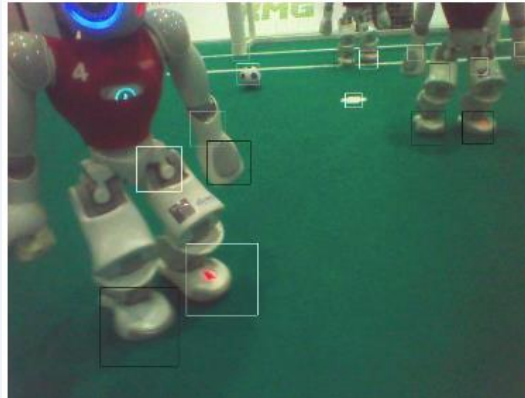
AGENDA

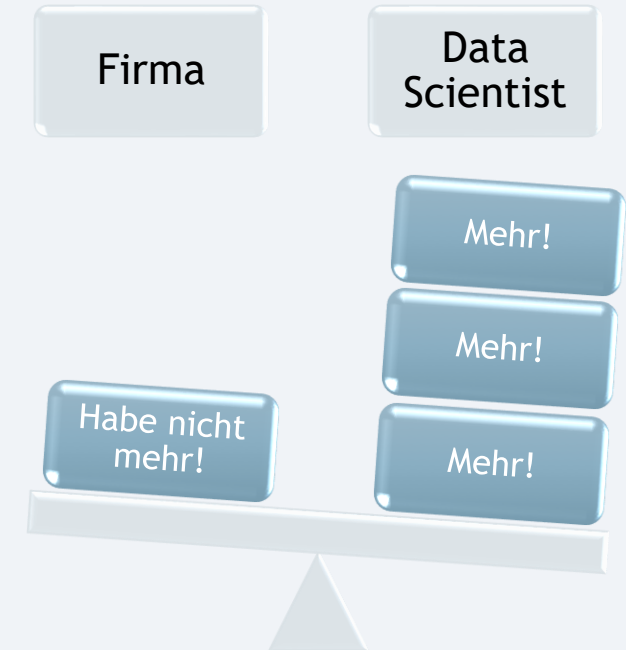
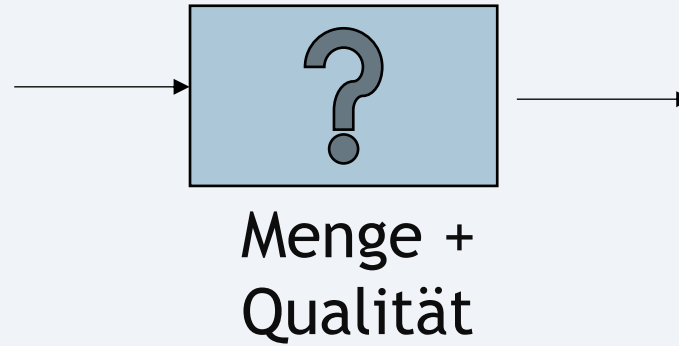
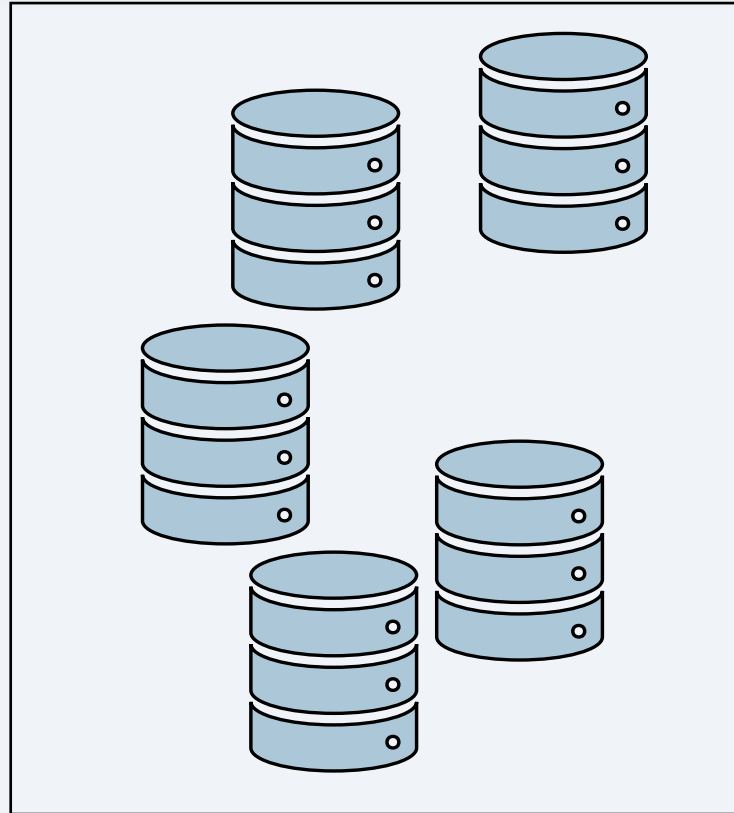
- Einführung
- Wie kommen Daten in die KI?
- Use Case Brainstorming
- Daten + AI-Act
- Fazit

MOTIVATION – BEST PRACTICE KI



MNIST 0-9	Kaggle A-Z
7 1 3 1 4 7 0 7 6 0	0 A A E N M S R Y O
2 2 3 1 5 4 6 1 1 9	S Q K B C B J O N V
8 2 2 4 9 8 3 6 2 7	M B S O T S T C N O
0 5 4 3 3 5 3 3 5 6	R P W m V O C D J X
1 6 3 0 7 4 4 7 7 6	F S S G B Z O O O U
3 9 5 2 7 4 6 4 4 4	E S S S R R R S O
1 5 6 8 4 7 7 0 3 5	4 I S T A W L L W Y
5 1 0 6 0 0 8 3 0 0	E W B S S S C S W W
4 0 2 6 4 6 4 4 1 1	O T m X O C E O S O
4 8 1 7 1 2 2 8 0 5	S Q Y Q E A P U S P

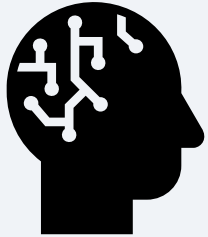




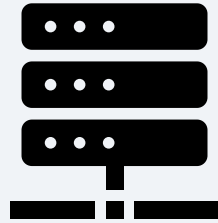


KURZES KENNENLERNEN

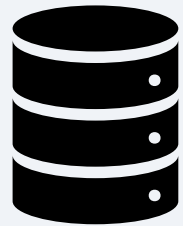
INITIALE PHASE



MA-Verfügbarkeit

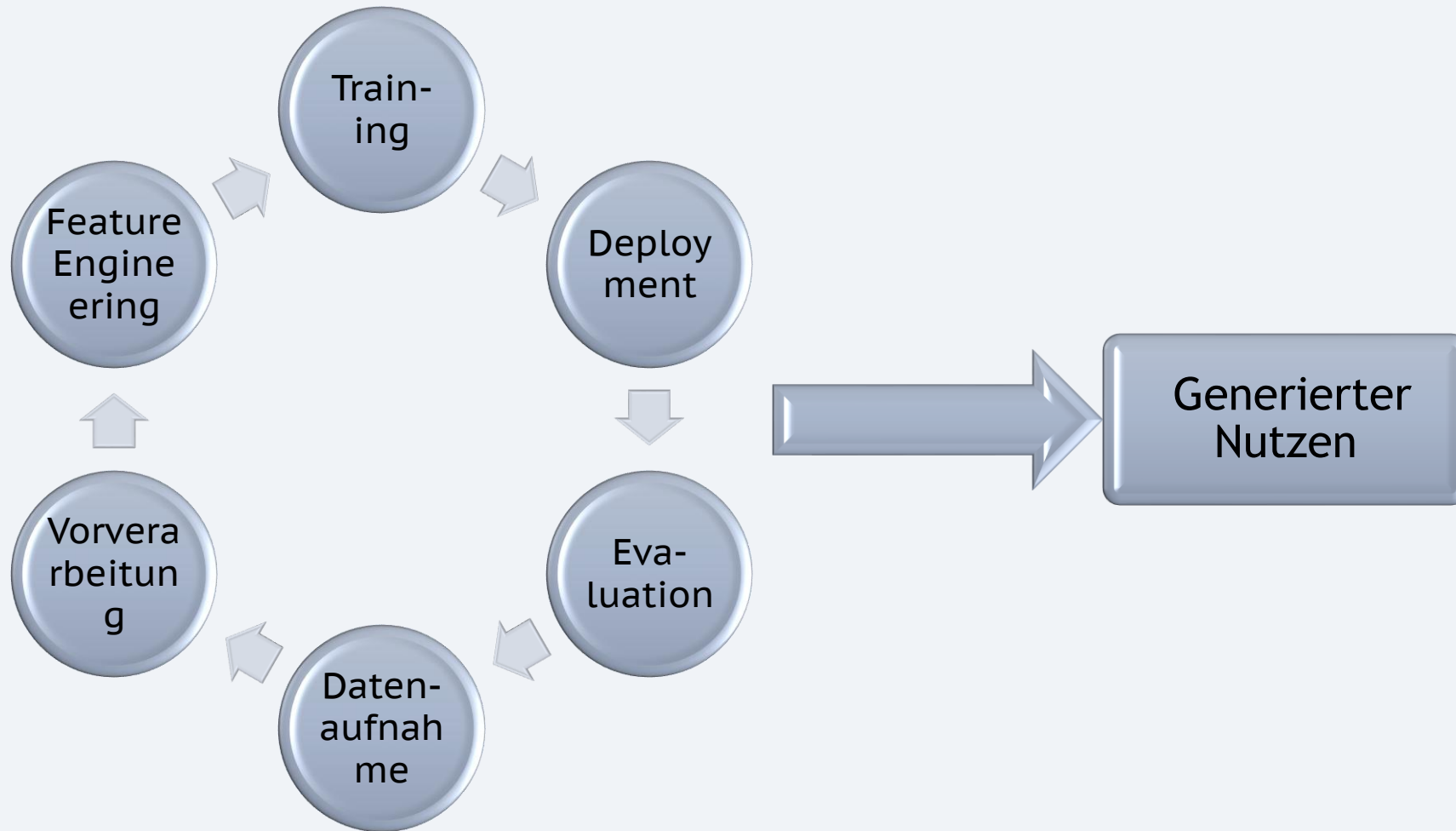


Rechenressourcen

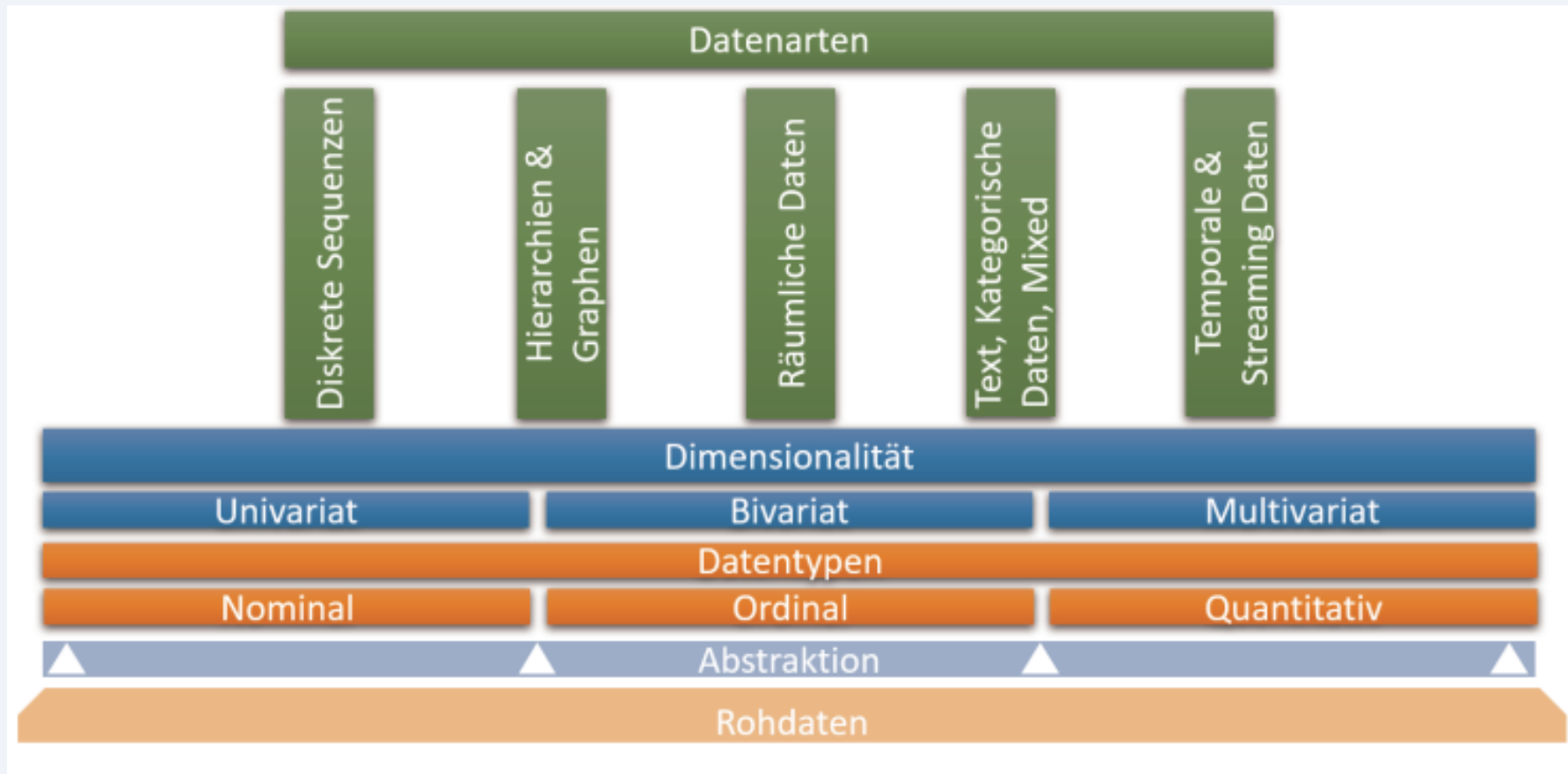


Datenlage

KI-PIPELINE

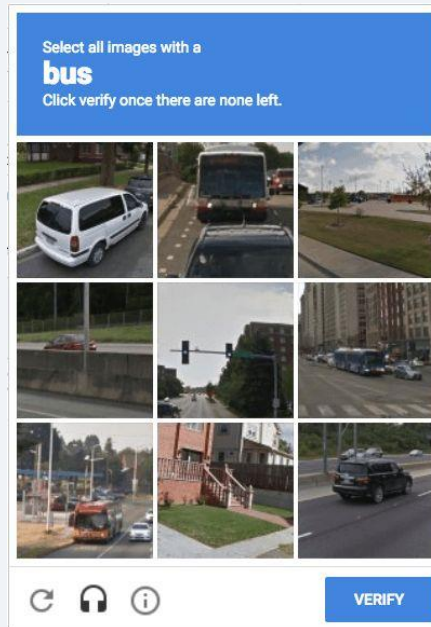


EINFÜHRUNG - DATENTYPEN



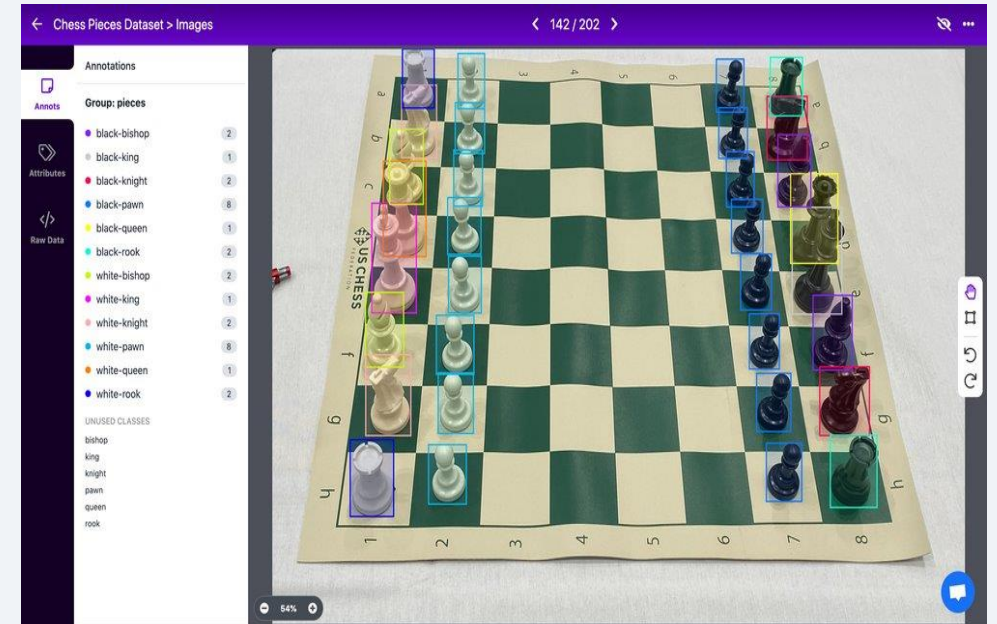
Nazemi, K., Kaupp, L., Burkhardt, D., & Below, N. (2021). 5.4
Datenvisualisierung. *De Gruyter Praxishandbuch Ser*, Walter de
Gruyter GmbH: Berlin, Germany, 477-502.

Klassifizierung:

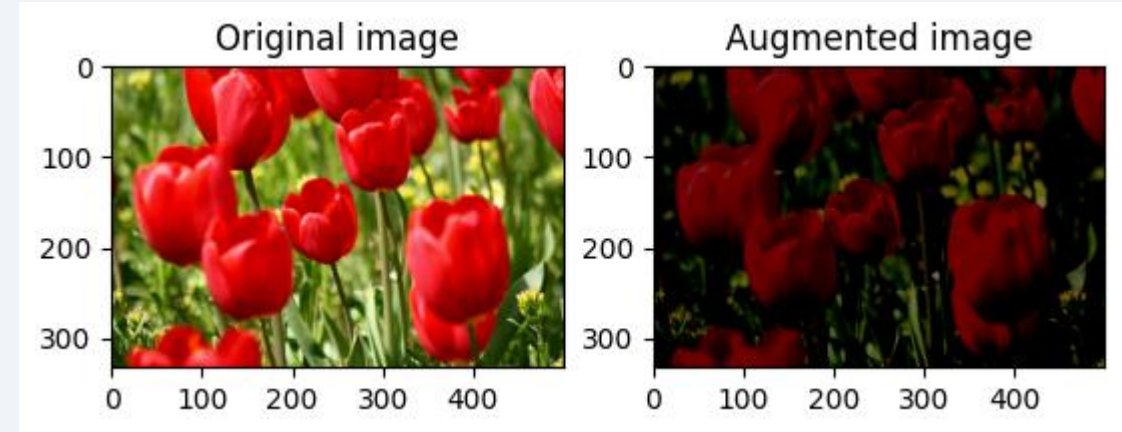
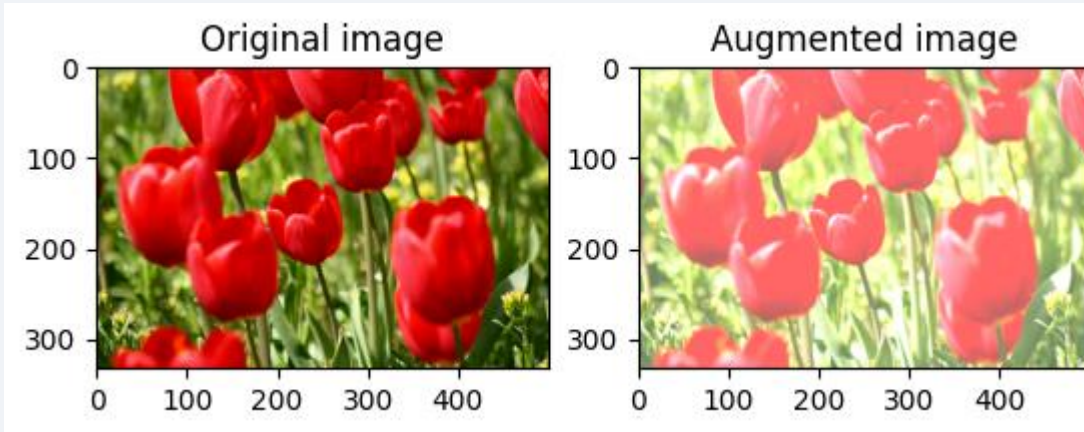
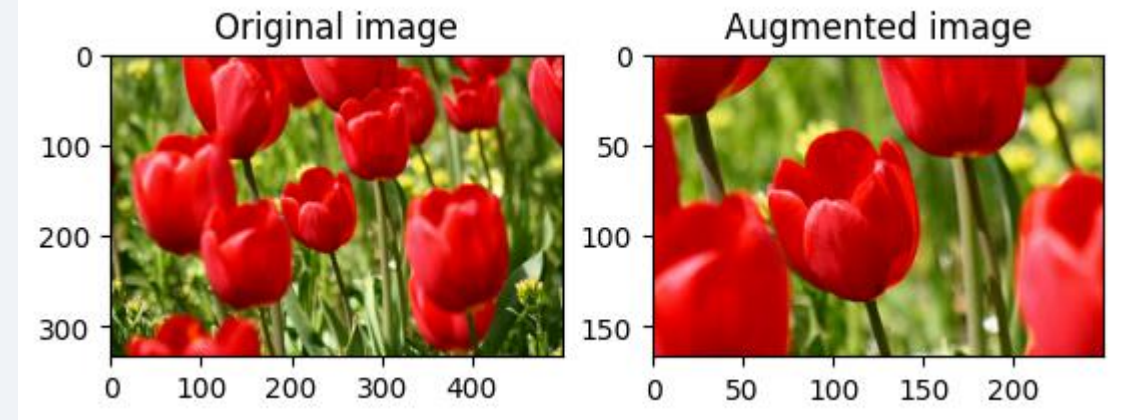
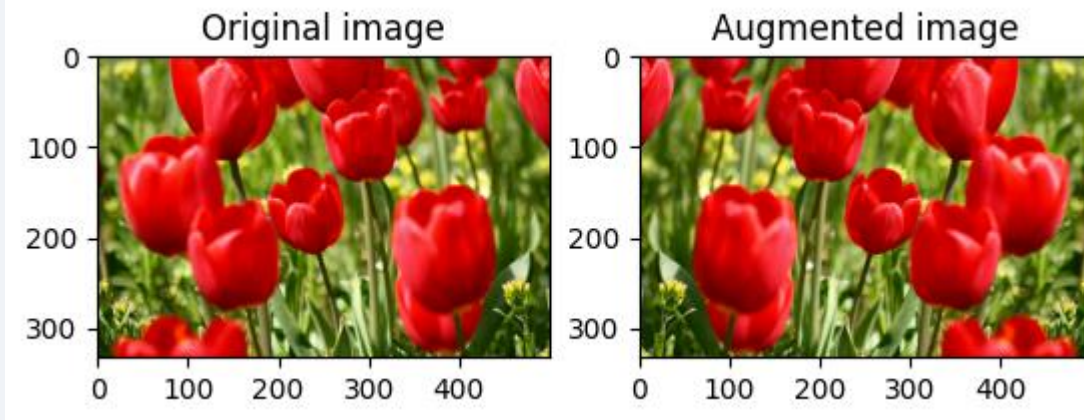


X0	X1	X2	Xn	y
0	0	0	0	1
0	1	1	0	1
0	1	0	1	0
1	1	1	0	0

Objekterkennung:

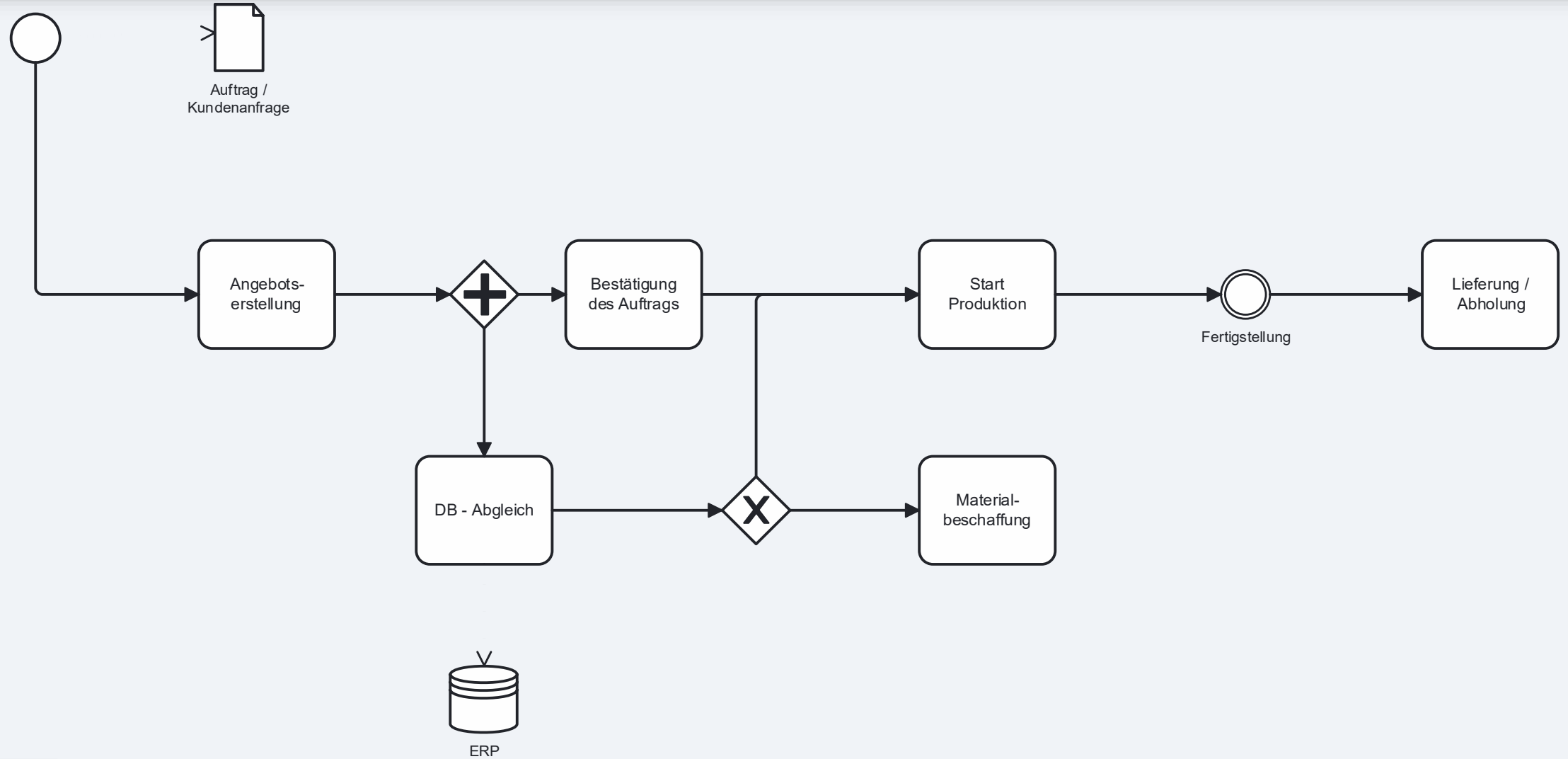


<https://blog.roboflow.com/tips-for-how-to-label-images/>

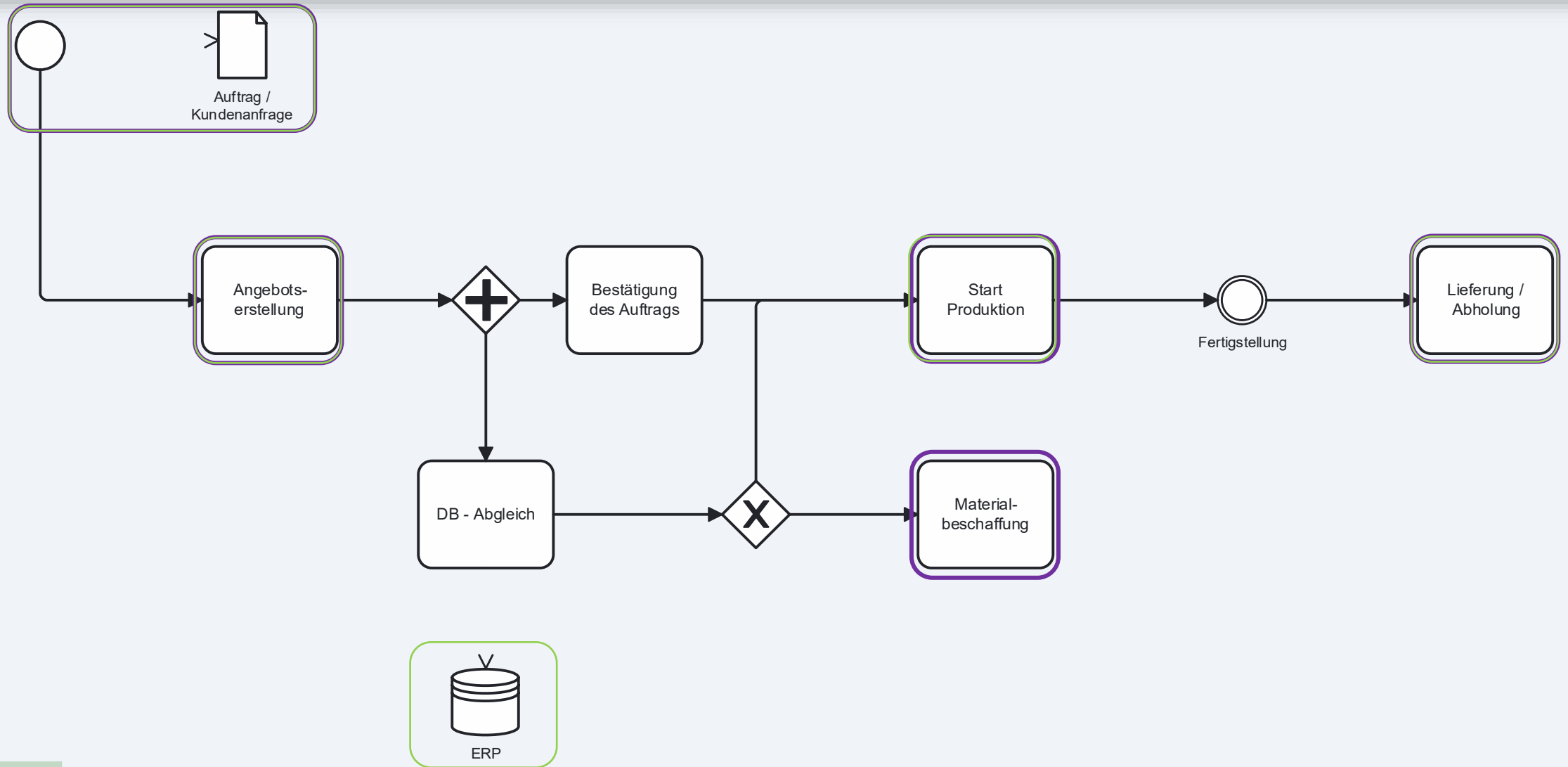


https://www.tensorflow.org/tutorials/images/data_augmentation

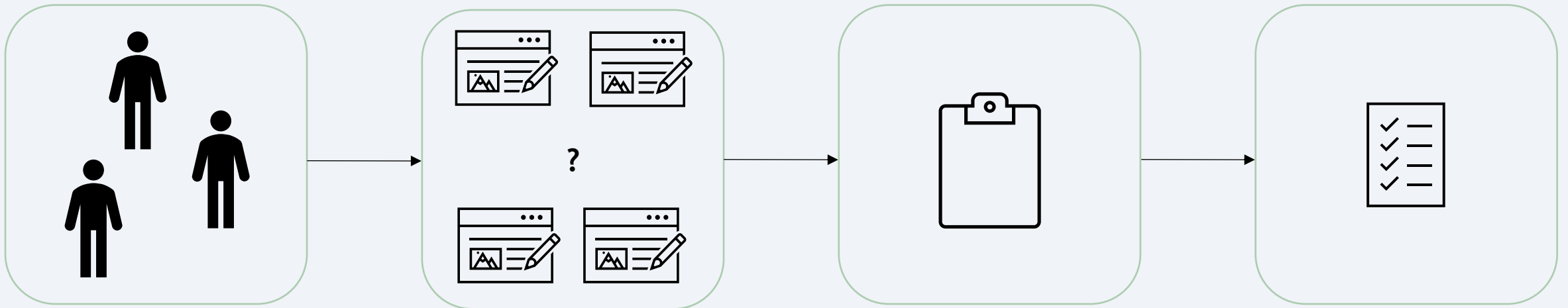
PROZESSMODEL: PRODUKTION



PROZESS – POINTS OF INTEREST



WER ÜBERNIMMT DEN AUFTRAG?



PIPELINE - BEISPIEL

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

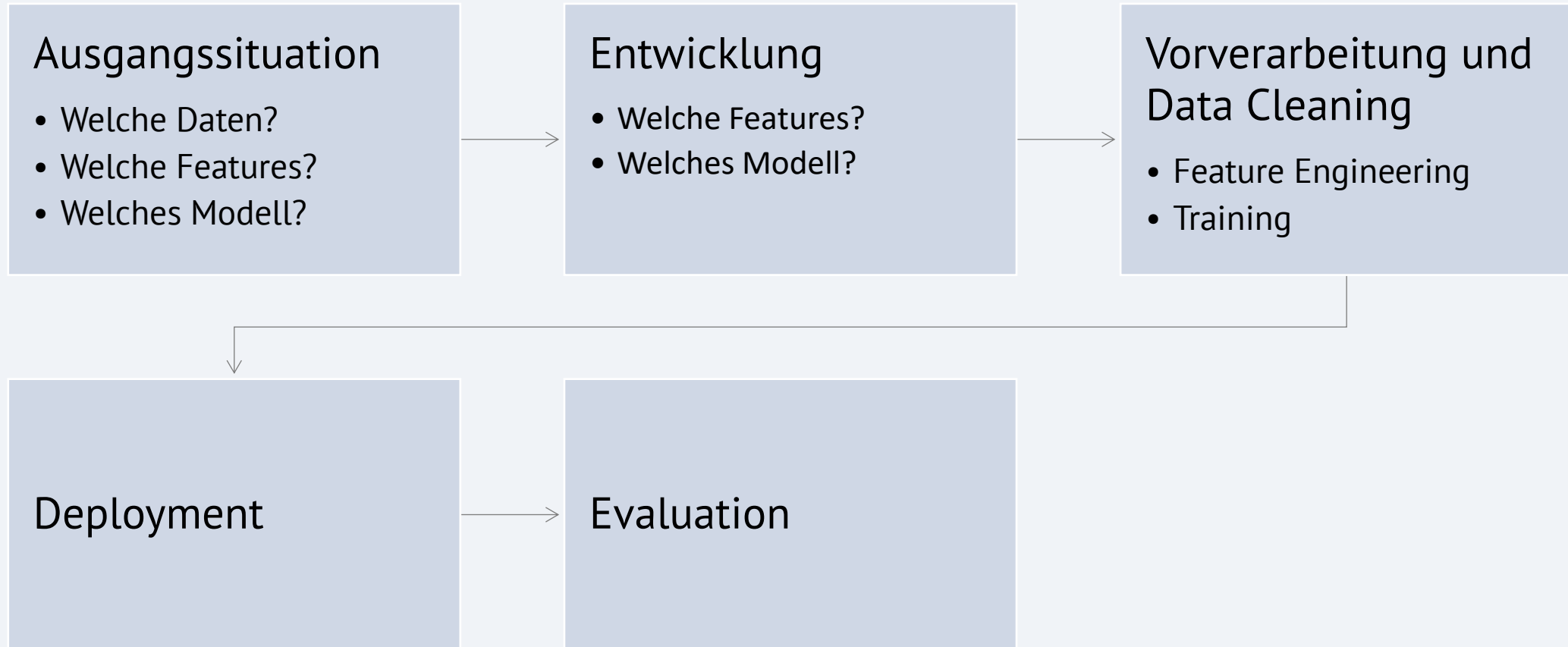
287

288

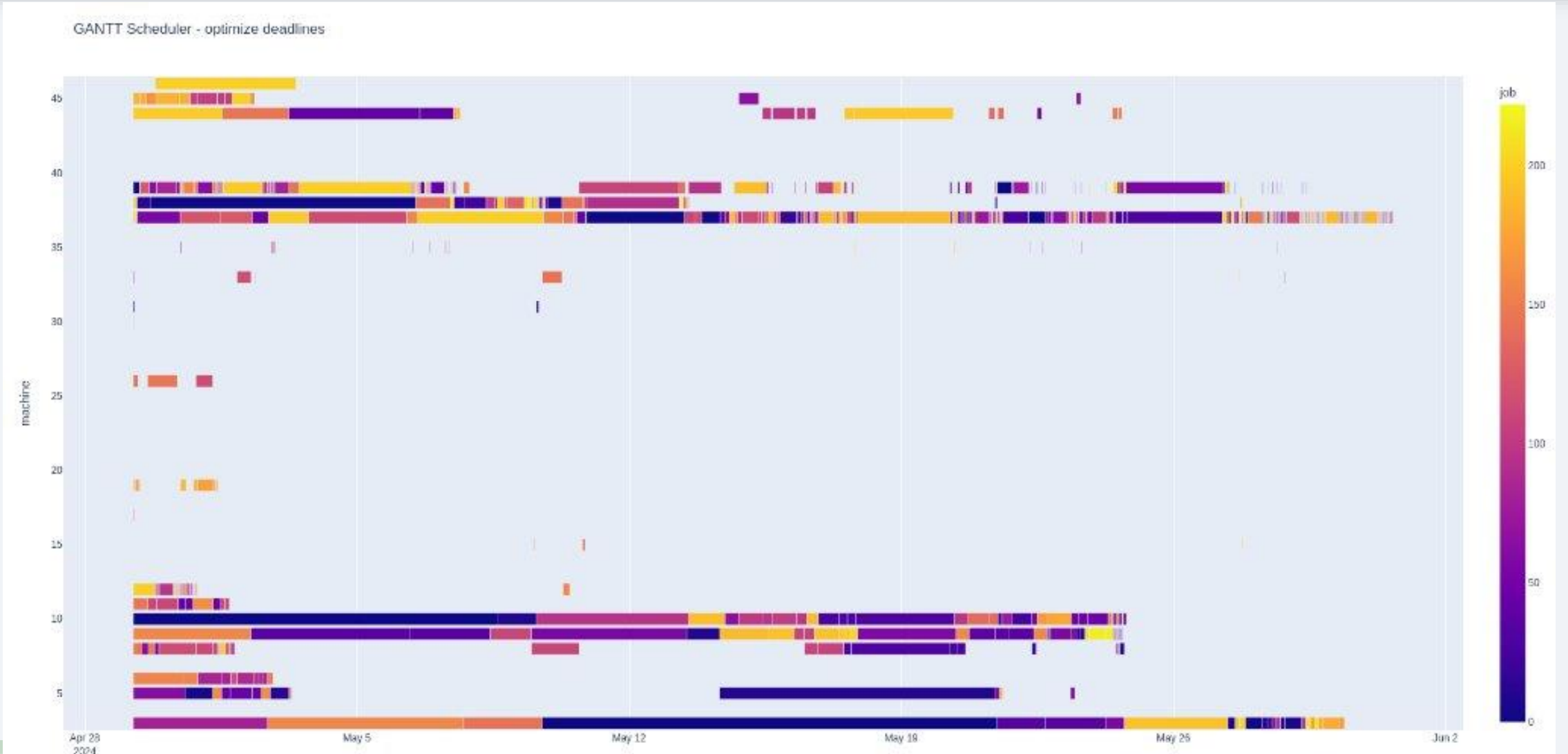
289

290

SCHEDULING I - PIPELINE



SCHEDULING II - ANWENDUNG



Datenqualität & -zugang



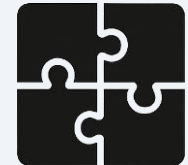
Datenschutz & Compliance



Fehlende Kompetenzen und Zeitmangel

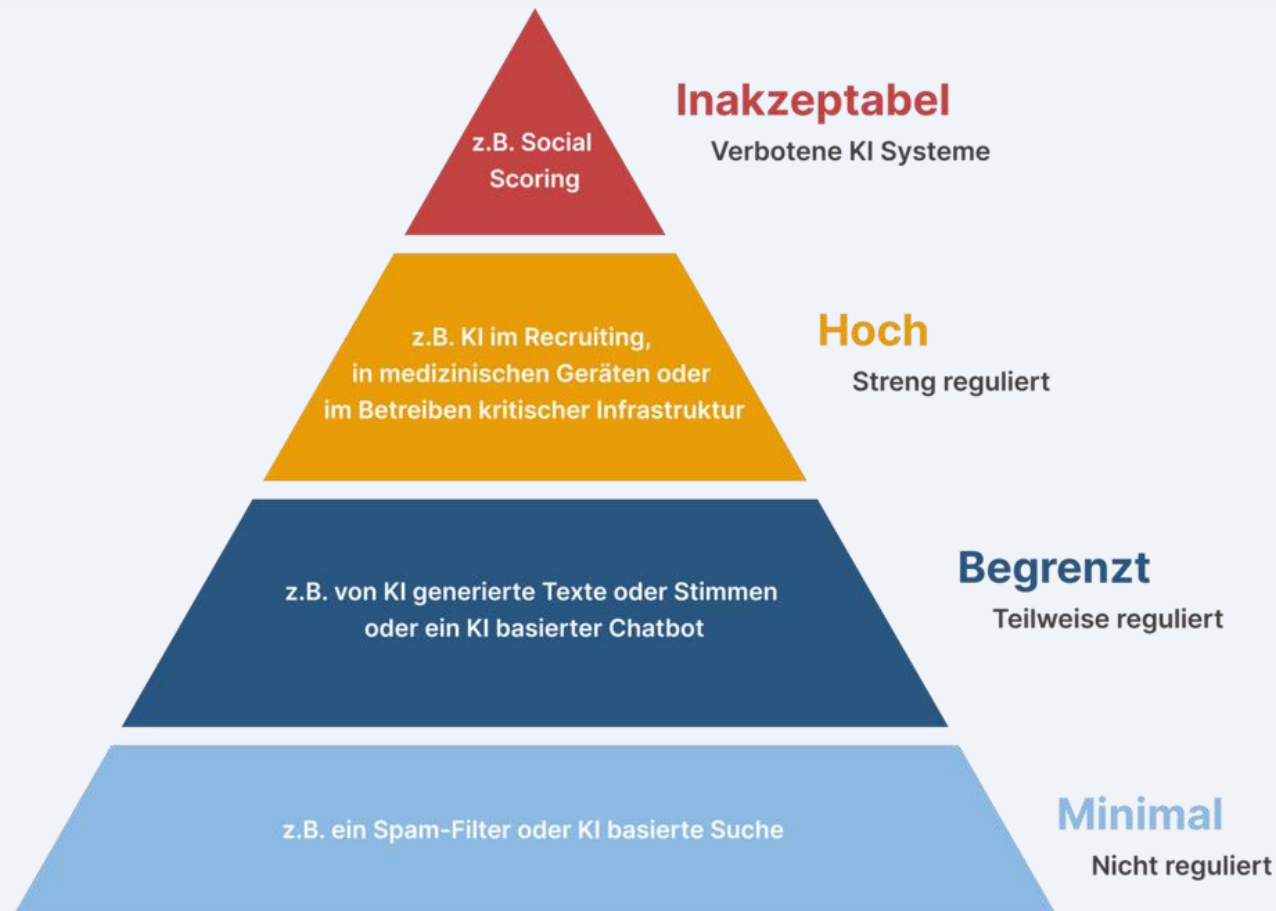


Integration / Heterogene Datenquellen

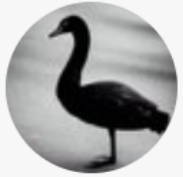


Ethische Bedenken + Bias + AI Act





<https://www.digitalhub-ai.de/de/nachricht-lesen/experten-artikel-der-eu-ai-act-und-seine-bedeutung-fuer-unternehmen>



Igor Babuschkin   @ibab · 16. Nov. 2024 

ChatGPT responses are contaminating every corner of the web, from scientific papers to Amazon reviews. OpenAI seems to be blissfully unaware of the issue.

 70

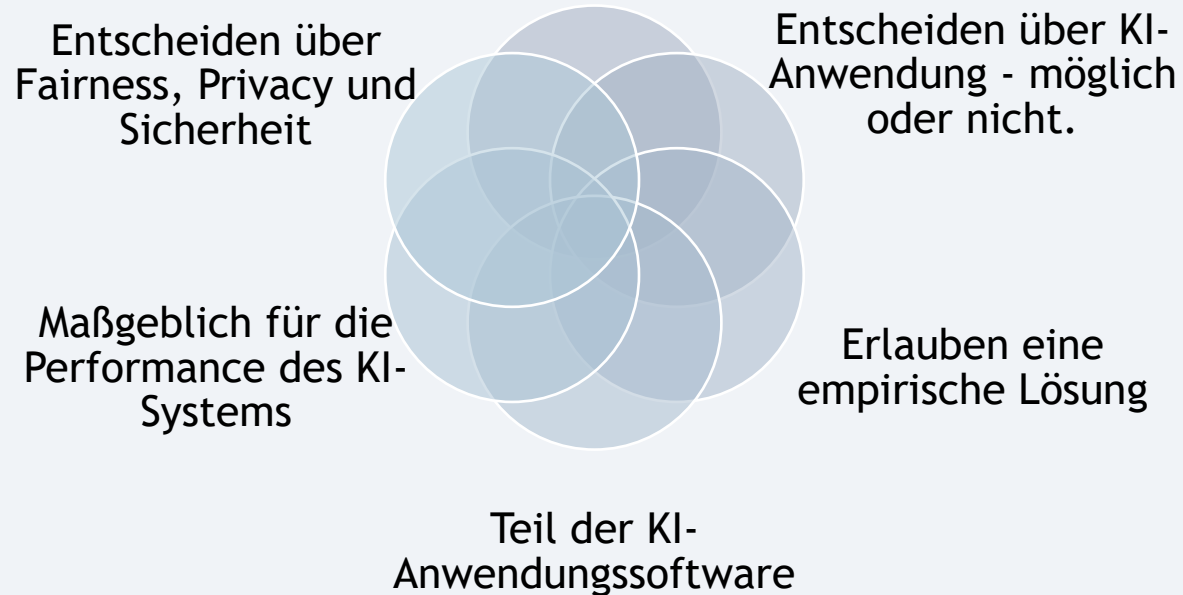
 204

 2.758

 111.583



Die Rolle der Daten



The importance of data

*"Data is the hardest part of ML and the most important piece to get right...
Broken data is the most common cause of problems in production ML systems"*
- [Scaling Machine Learning at Uber with Michelangelo](#) - Uber

"No other activity in the machine learning life cycle has a higher return on investment than improving the data a model has access to."
- [Feast: Bridging ML Models and Data](#) - Gojek

VIELEN DANK!

KONTAKT

www.kmi-leipzig.de
www.kmi-netzwerk.org

Email:

steven.lehmann@infai.org

schliesser@infai.org



Das Forschungs- und Entwicklungsprojekt KMI wird im Rahmen der Fördermaßnahme „Zukunft der Arbeit: Regionale Kompetenzzentren der Arbeitsforschung – Künstliche Intelligenz“ im Programm „Innovationen für die Produktion, Dienstleistung und Arbeit von morgen“ des Bundesministeriums für Forschung, Technologie und Raumfahrt (BMFTR) gefördert und vom Projektträger Karlsruhe (PTKA) betreut.

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

