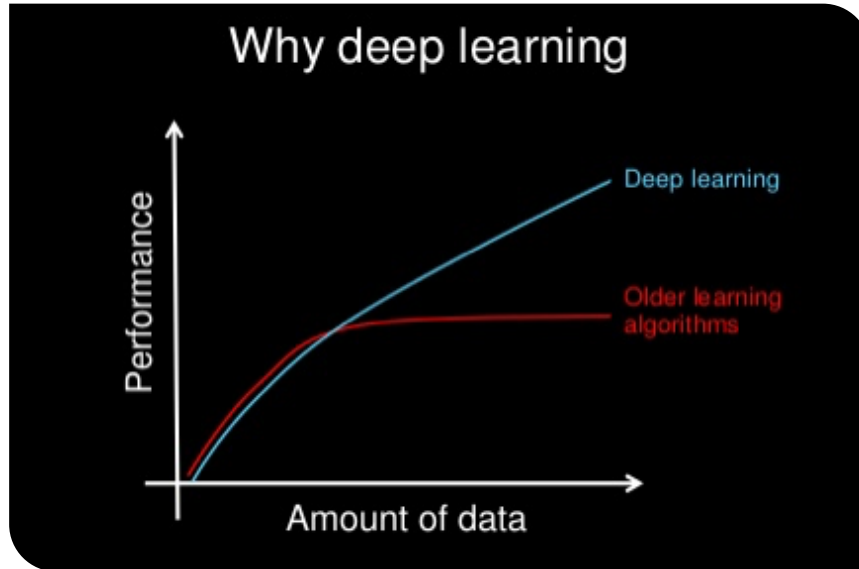


Label Convergence with Interactive Machine Learning for Brand Identification in Print Advertisements

**Interactive Analytics Seminar Topic
Merlin Knaeble – Summer Term 2021**

Machine Learning Needs Training Data



BaiDu Research Keynote (Ng, 2015)

Proven scaling past
300 million
training images!

Sun et al. (2017)

Training Data Requires Manual Labeling

“costly, error-prone, labor-intensive activity” (Nadj et al. 2020, p. 131)



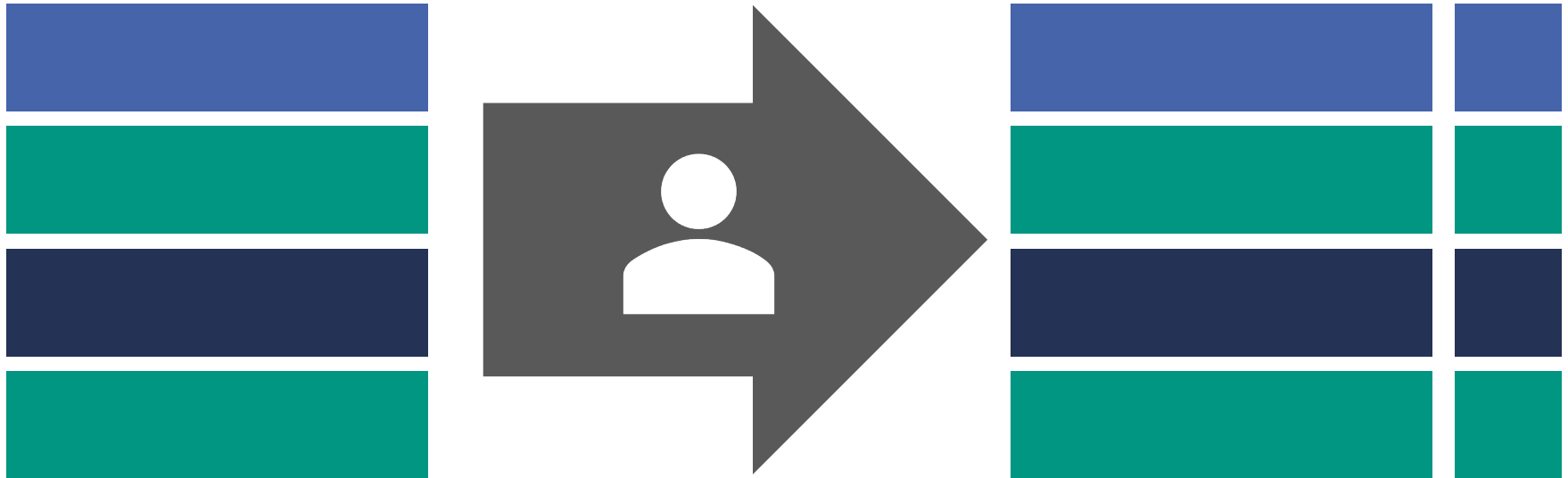
“I’m doing labeling for the money.”



“boring, repetitive, never-ending work”

Workers in Data Factories (New York Times 2018; BBC 2018)

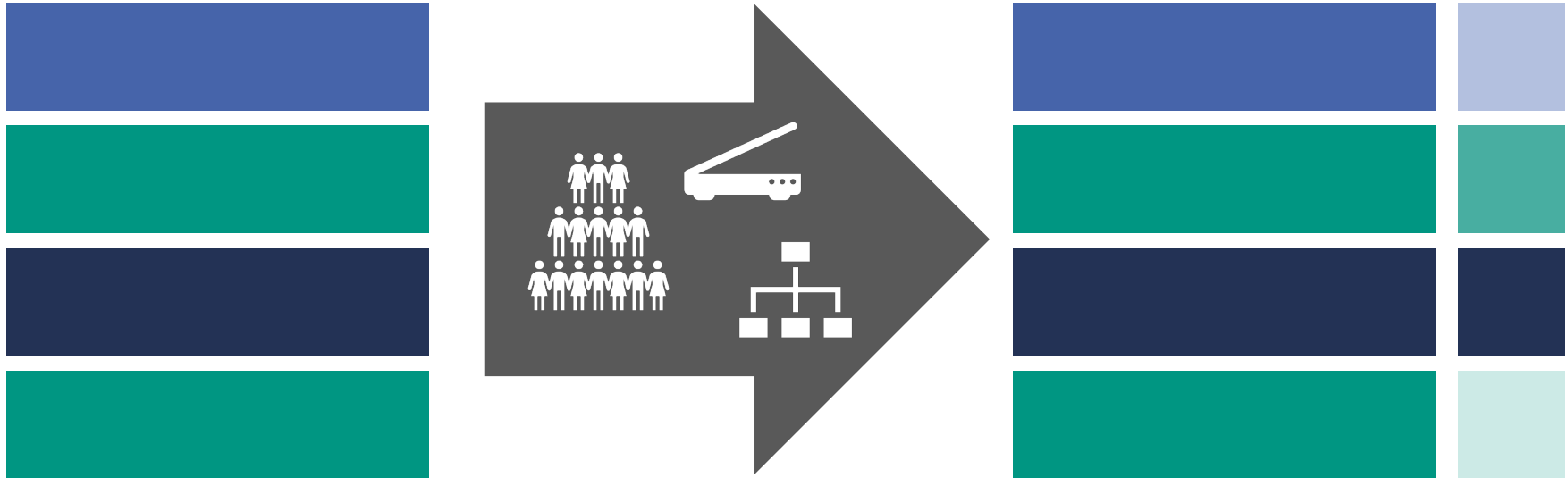
Definition of Labeling



information enclosure to objects (Bernard et al. 2018)

classes, bounding boxes, segmentations, ...

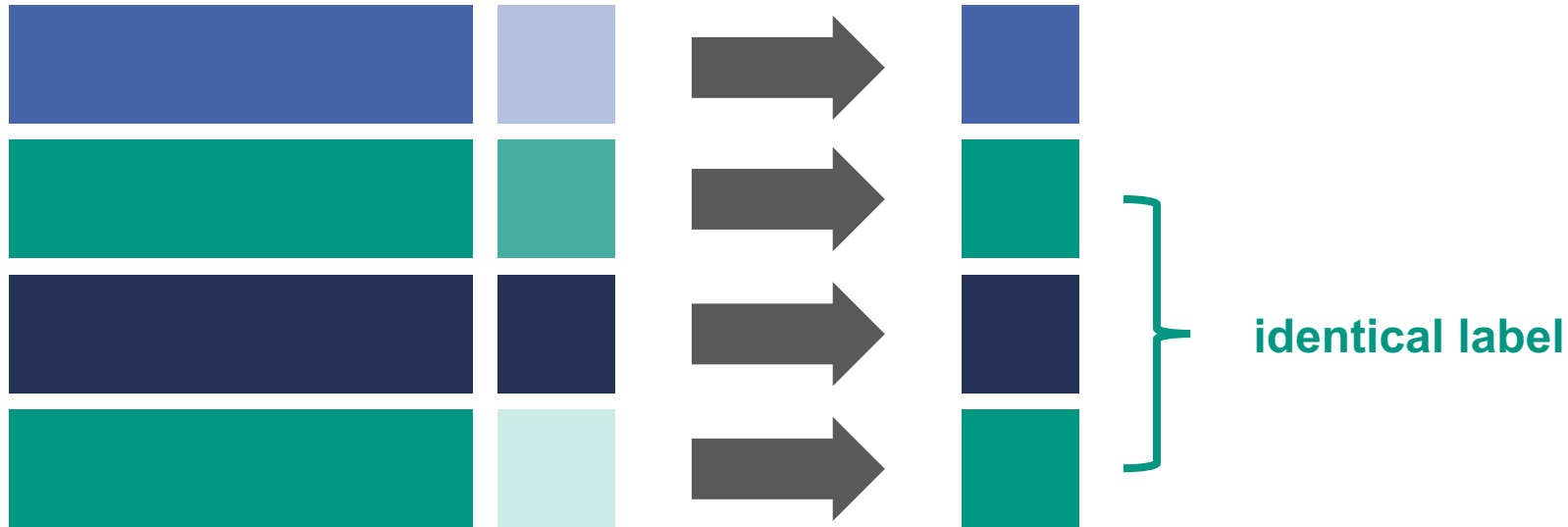
Weak & Ambiguous Labels



coordination issues in crowdsourcing
different semantic understanding of topics
automatic data extraction (e.g. OCR)
...

**e.g. when using
free-text labels**

Label Convergence



The original labels were not wrong!
Just inaccurate & not comparable...



Project Context

Context of the Project

Funded by



Deutsche
Forschungsgemeinschaft

German Research Foundation

- “Advertisement over Time: Interactive Machine Learning for Image Data Analysis” → DFG funded research project
- joint project with University of Mannheim
- two major components (marketing/ads & interactive machine learning) centered around common data-set

Available Data

Data Set
over 300k images
1843 – today
The Economist



Meta-Data

General

Date
Volume/Issue
Size
Advertiser

Objects

Type
Number
Position
Size

People

Age
Gender
Ethnicity
Number
Position
Size

Colors

Dominant
Colors
Contrasts





Your Task

Focus on the Brand Name

Data Set
over 300k images
1843 – today
The Economist



Meta-Data

General

Date
Volume/Issue
Size

Advertiser

Objects

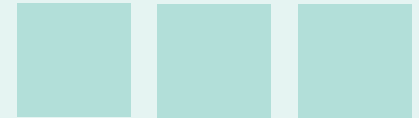
Type
Number
Position
Size

People

Age
Gender
Ethnicity
Number
Position
Size

Colors

Dominant
Colors
Contrasts



Advertiser/Brand Names as Labels are Volatile

Introducing the CLA-Class.

DaimlerChrysler

Mercedes

Mercedes Benz

Mercedes-AMG

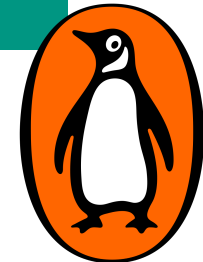
Mercedes-
Benz AG

Mercedes-Benz

The best or nothing.
Mercedes-Benz

Mercedes-
Benz
USA, LLC

The new Mercedes-Benz C-Class!



former subsidiary
"HPBooks"

Interactive Analytics Seminar

Work Packages

- condensing the current list of brands into a usable range
- **developing analytics prototype**
 - **unsupervised ML approaches**
 - **string distance metrics**
 - **user interface to include experts**
- coordinating with marketing experts from Uni Mannheim for a final evaluation of your work

Requirements

- practical experience and/or theoretical knowledge on unsupervised ML and string distance metrics
- experience with the technologies for ML, front- & backend



Merlin Knaeble

Doctoral Researcher
Interactive Machine Learning

merlin.knaeble@kit.edu

