# Conclusions and Future Directions

Overview and Introduction

Knowledge Extraction

Knowledge Cleaning

Q&A

Break

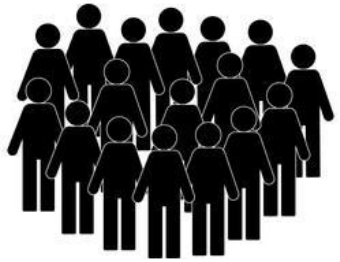
Ontology Mining

Applications

**Conclusion and Future Directions** **10 min**

Q&A

# An End-to-End Pipeline



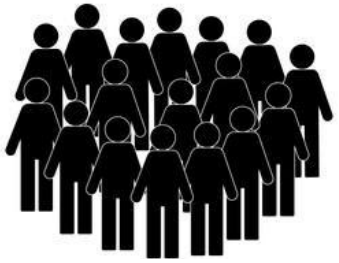
Understand domain and attributes, and generate LOTS OF training data

Identify product taxonomy and attributes

Train and fine-tune models

Extraction Models

Postprocess extraction results to further improve data quality



Pre-publish evaluation as gatekeeper to guarantee high quality data

# An End-to-End Pipeline

Automatic Training Data Generation

Distant supervision, Data programming

Automatic Taxonomy Enrichment

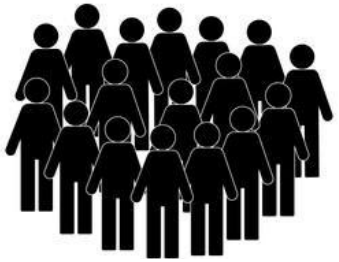
Category classification, attribute identification

Train and fine-tune models

Extraction Models

Benchmarking

Postprocess extraction results to further improve data quality



Pre-publish evaluation as gatekeeper to guarantee high quality data

# An End-to-End Pipeline

Automatic Training Data Generation

Distant supervision, Data programming

Automatic Taxonomy Enrichment

Category classification, attribute identification

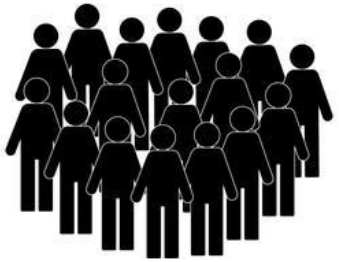
Train and fine-tune models

Extraction Models

Benchmarking

Deep Learning Data Cleaning

Postprocess extraction results to further improve data quality



Pre-publish evaluation as gatekeeper to guarantee high quality data

# An End-to-End Pipeline

Automatic Training Data Generation

Distant supervision, Data programming

Automatic Taxonomy Enrichment

Category classification, attribute identification

Train and fine-tune models

Extraction Models

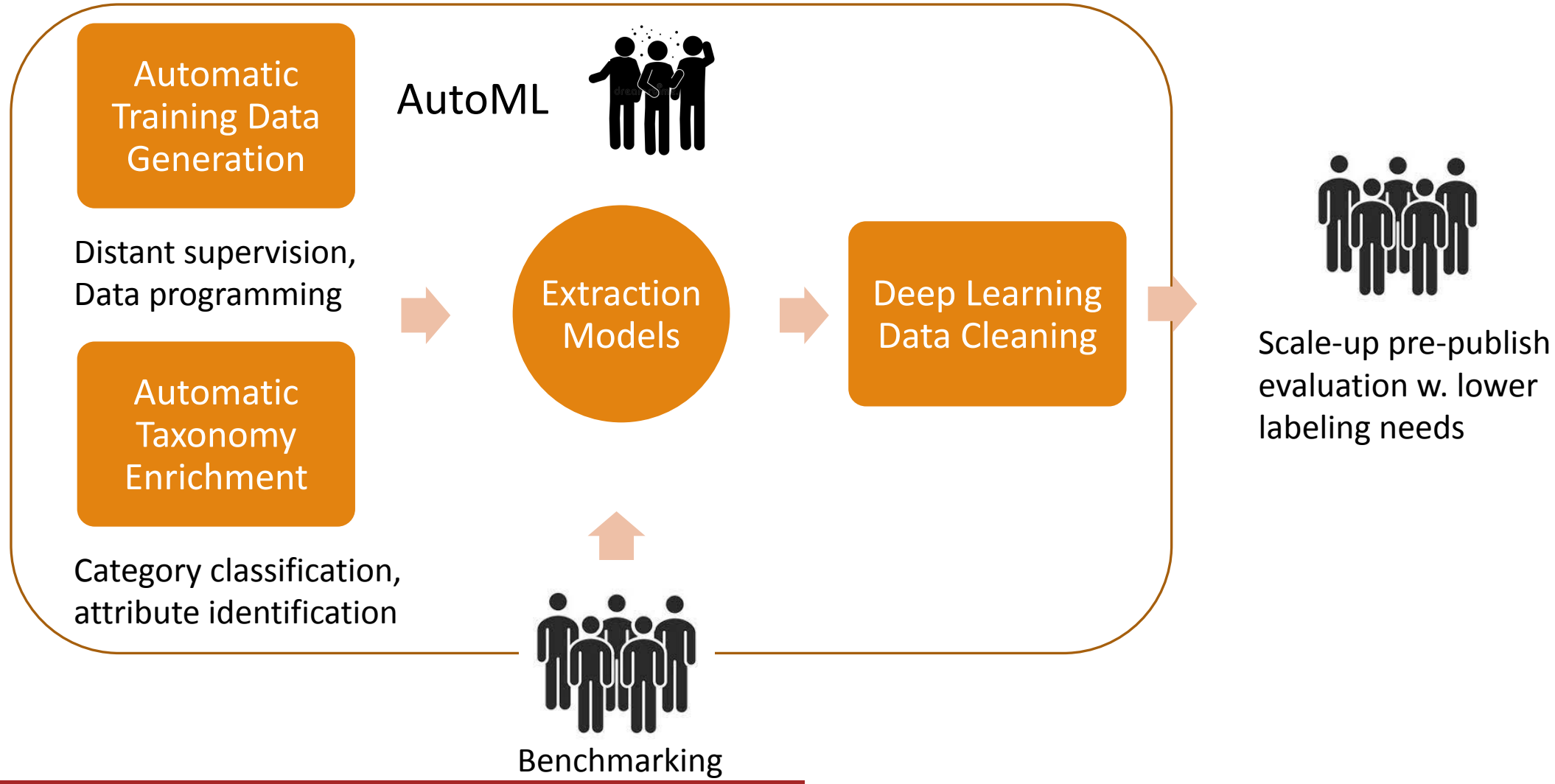
Benchmarking

Deep Learning Data Cleaning

Postprocess extraction results to further improve data quality

Scale-up pre-publish evaluation w. lower labeling needs

# An End-to-End Pipeline



# Knowledge Extraction Takeaways

- Modeling attribute value prediction as Sequence Tagging, Question Answering and Text generation task.

- Using the attribute name embedding and product type taxonomy embedding attend to text profile.
  - Improve the performance.
  - Generalizability on few-shot/zero-shot learning.

- Opportunities in combining text, text on  image, image feature by utilizing multi-modal transformer to allow interaction between all features.

# Knowledge Cleaning Takeaways

- Knowledge cleaning essentially detects inconsistency in data within attribute, across different attributes, and across different data sources.
- All data cleaning methods complement each other and effective ensemble them can maximize the final performance
- All these techniques are generic and applicable to KGs in other domains

# Ontology Mining Takeaways

- Ontology mining discovers
  - Emerging product categories and attributes,
  - The relations between product categories a
  - The relations between product categories and attributes.

- Training data can be scarce and noisy. Supervision from data itself and customer behavior signals are very useful when leveraged properly.

- The presented techniques are applicable for KGs in other domains

# Applications Takeaways

- Applications of product knowledge graphs can make use of:
  - The structured factual information for each product.
  - The product connections in the overall graph structure.

- The graph structure also allows the utilization of graph level constructs, like knowledge graph embeddings, which is useful for many applications.

- General applications of knowledge graphs include question answering, recommendation, search, among others.

# Questions

Overview and Introduction

Knowledge Extraction

Knowledge Cleaning

Q&A

Break

Ontology Mining

Applications

Conclusion and Future Directions

**Q&A**