Sentiment analysis et al.: Machine Learning and NLP

- Approaches to sentiment analysis today:
    - Machine-learning statistical approaches mostly supervised
    - Natural Language must be converted into numbers!
    - Support from linguistic theories and cognitive theories: philosophy of language, cognitive science, psychology
- Supervised text classification
    - Training: text + labels (examples of correct classifications) → model
        - Finding patterns, regularities, features!
    - Prediction: text + model → labeled text
        - Testing the system with NEW examples: comparing the results with the data annotated by humans
- New tendencies:
    - Neural network models and deep learning, exploitation of Large Language Models, LLMs
- Training and testing the accuracy of automatic systems in NLP requires the availability of
    - Corpora annotated by humans (gold truth)
    - Benchmark and Evaluation Campaigns

Organization of Shared Tasks

Corpora as benchmarks for novel states of the art

Evalita: https://www.evalita.it/campaigns/evalita-2023/

Semeval: https://semeval.github.io/

EVALITA
Evaluation of NLP and Speech Tools for Italian

Example: the Sentipolc dataset

Table 1: Combinations of values allowed by our annotation scheme

`subj`	`pos`	`neg`	`iro`	**description**
0	0	0	0	an objective tweet example: *l'articolo di Roberto Ciccarelli dal manifesto di oggi* `http://fb.me/1BQVy5WAk`
1	0	0	0	a subjective tweet with neutral polarity and no irony example: *Primo passaggio alla #strabrollo ma secondo me non era un iscritto*
1	1	0	0	a subjective tweet with positive polarity and no irony example: *splendida foto di Fabrizio, pluri cliccata nei siti internazionali di Photo Natura* `http://t.co/GWoZqbxAuS`
1	0	1	0	a subjective tweet with negative polarity and no irony example: *Monti, ripensaci: l'inutile Torino-Lione inguaia l'Italia: Tav, appello a Mario Monti da Mercalli, Cicconi, Pont...* `http://t.co/3CazKS7Y`
1	1	1	0	a subjective tweet with positive and negative polarity (mixed polarity) and no irony example: *Dati negativi da Confindustria che spera nel nuovo governo Monti. Castiglione: "Avanti con le riforme"* `http://t.co/kIKnbFY7`
1	1	0	1	a subjective tweet with positive polarity, and an ironic twist example: *Letta: sicuramente non farò parte del governo Monti . e siamo un passo avanti. #finecorsa*
1	0	1	1	a subjective tweet with negative polarity, and an ironic twist example: *Botta di ottimismo a #lInfedele: Governo Monti, o la va o la spacca.*

sentipolc @ evalita

SENTIment POLarity Classification task

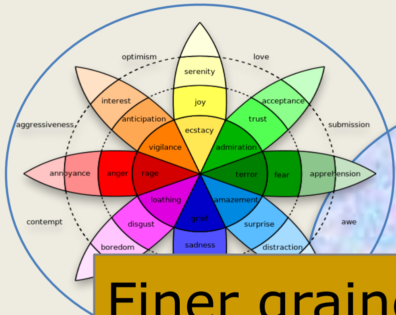
call for participation

Evaluation campaigns

- Semeval: (English, Multilingual): https://semeval.github.io/
- EVALITA (Italian): https://www.evalita.it/
- IberEval, IberLEF (Spanish and other Iberian languages): https://sites.google.com/view/iberlef-2024/calls?authuser=0
- GermEval (German): https://germeval.github.io/
- You! : https://docs.google.com/document/d/1yddh0mePk7-CNYoT7lA2GgowuOVhXvX664TSMq4MIRM/edit?usp=sharing

Computational models and lexical resources for affect

- Different facets of the affective content
- Wide availability of lexical resources for English covering the various perspectives.
- Both sentiment and emotion lexicons, and psycholinguistic resources available for English, refer to various affective models and capture different facets of affect, including:
  - Sentiment polarity: aspects related to the polarity of words
  - Finer-grained aspects: which can be captured according to different categorical or dimensional models of emotions.



Sentiment-related features

Finer grained affective features

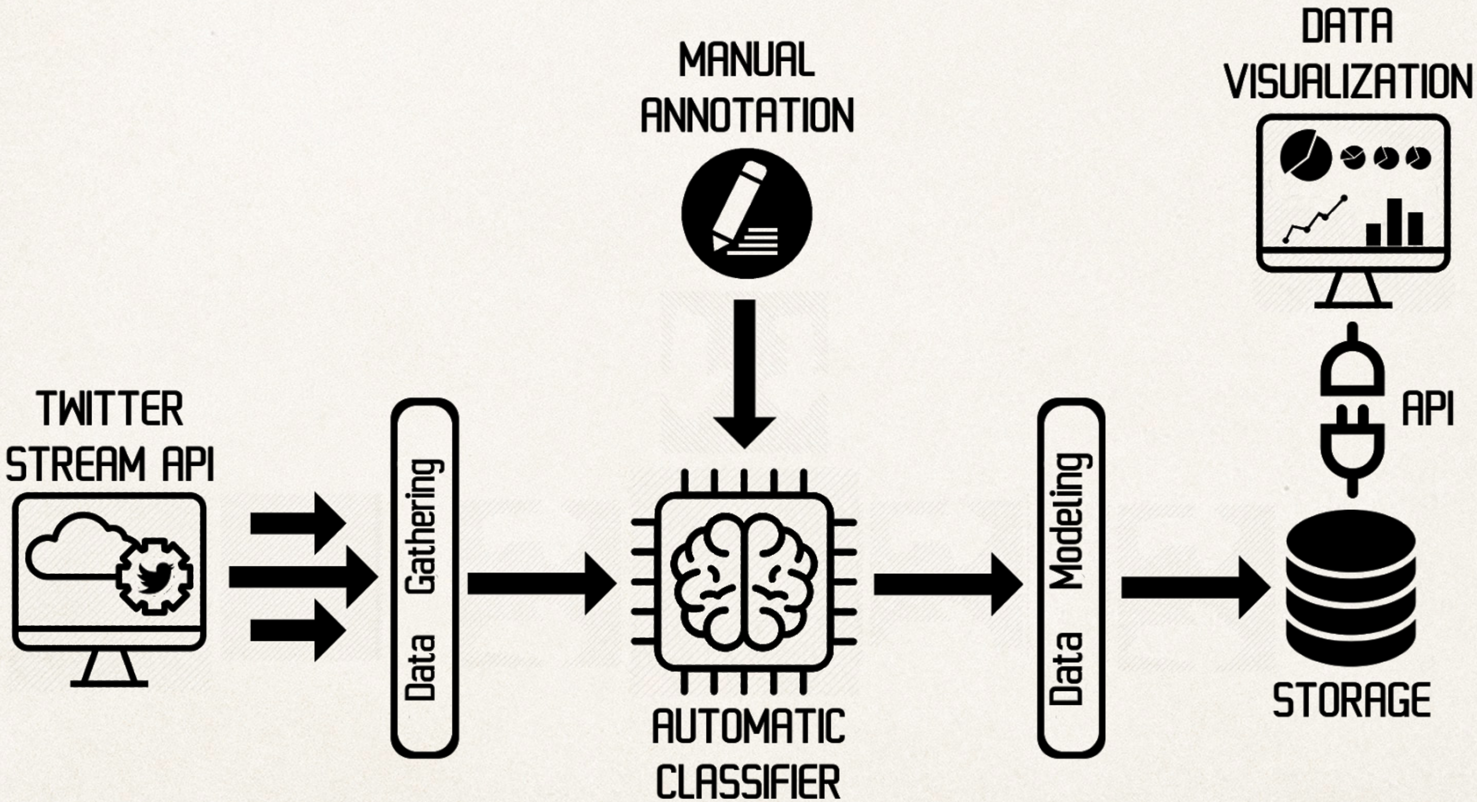
Combining different analysis and technologies

- Detecting and visualizing hate speech
- Contro l'odio: tecnologie informatiche, percorsi formativi e storytelling partecipativo per combattere l'intolleranza
    - Hate maps: https://mappa.controlodio.it/
    - Target: immigrants

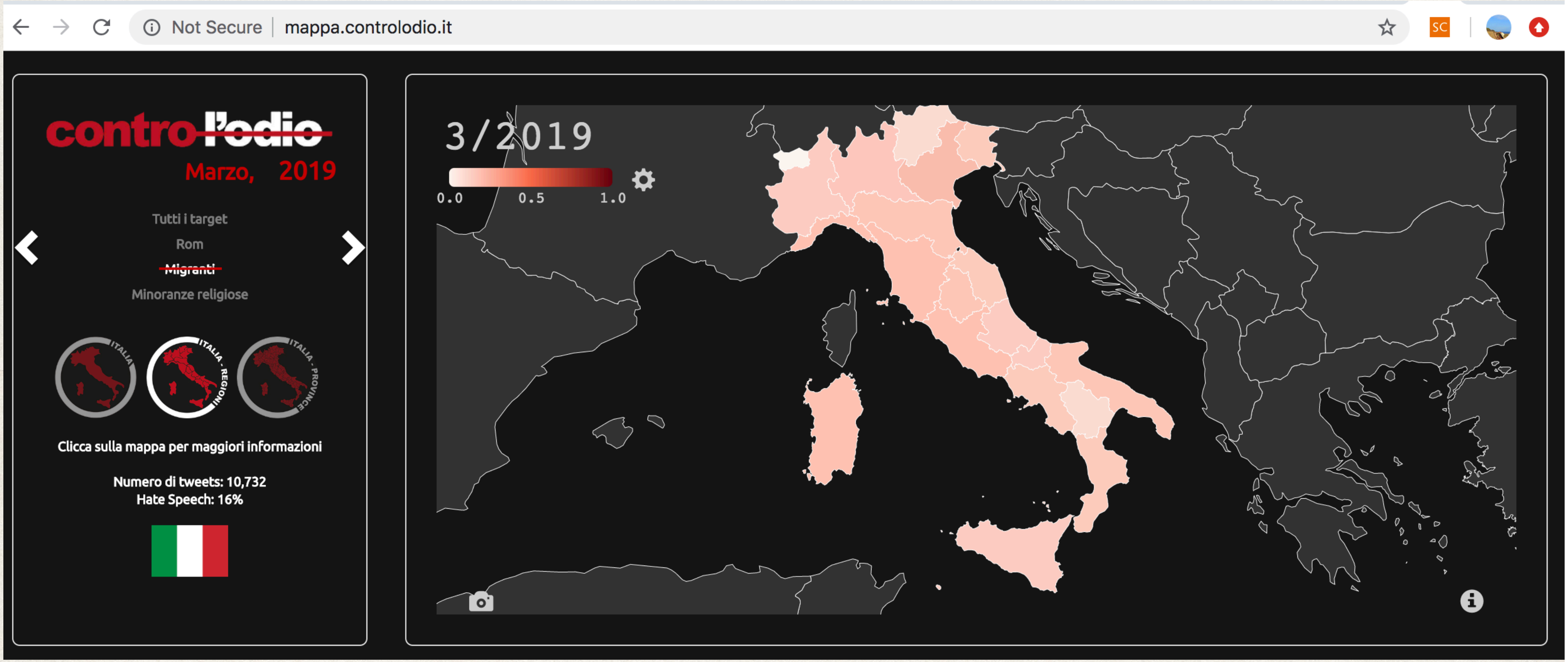
MINISTERO del LAVORO e delle POLITICHE SOCIALI
DIREZIONE GENERALE DEL TERZO SETTORE E DELLA RESPONSABILITÀ SOCIALE DELLE IMPRESE

- Natural Language Processing:
  Development of linguistic resources
  Models for monitoring (robustness w.r.t. time)
- Data Analysis applied to social media for detecting the dynamics for the diffusion of the HS
- Data Visualization: interactive visualization of complex information for allowing the access to data previously collected and analyzed
- Educational tools

Possible pipeline

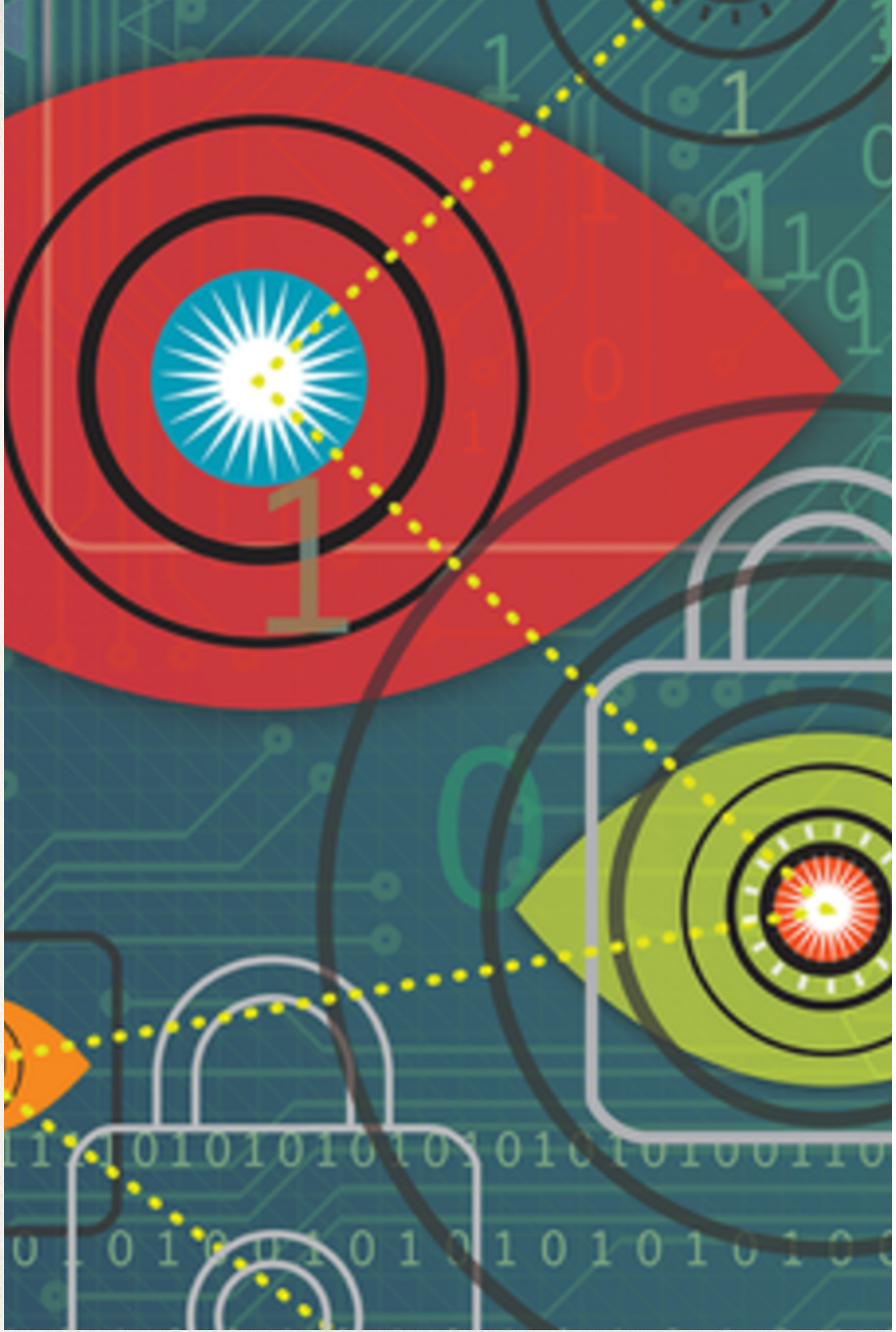


Detecting and Visualizing hate speech



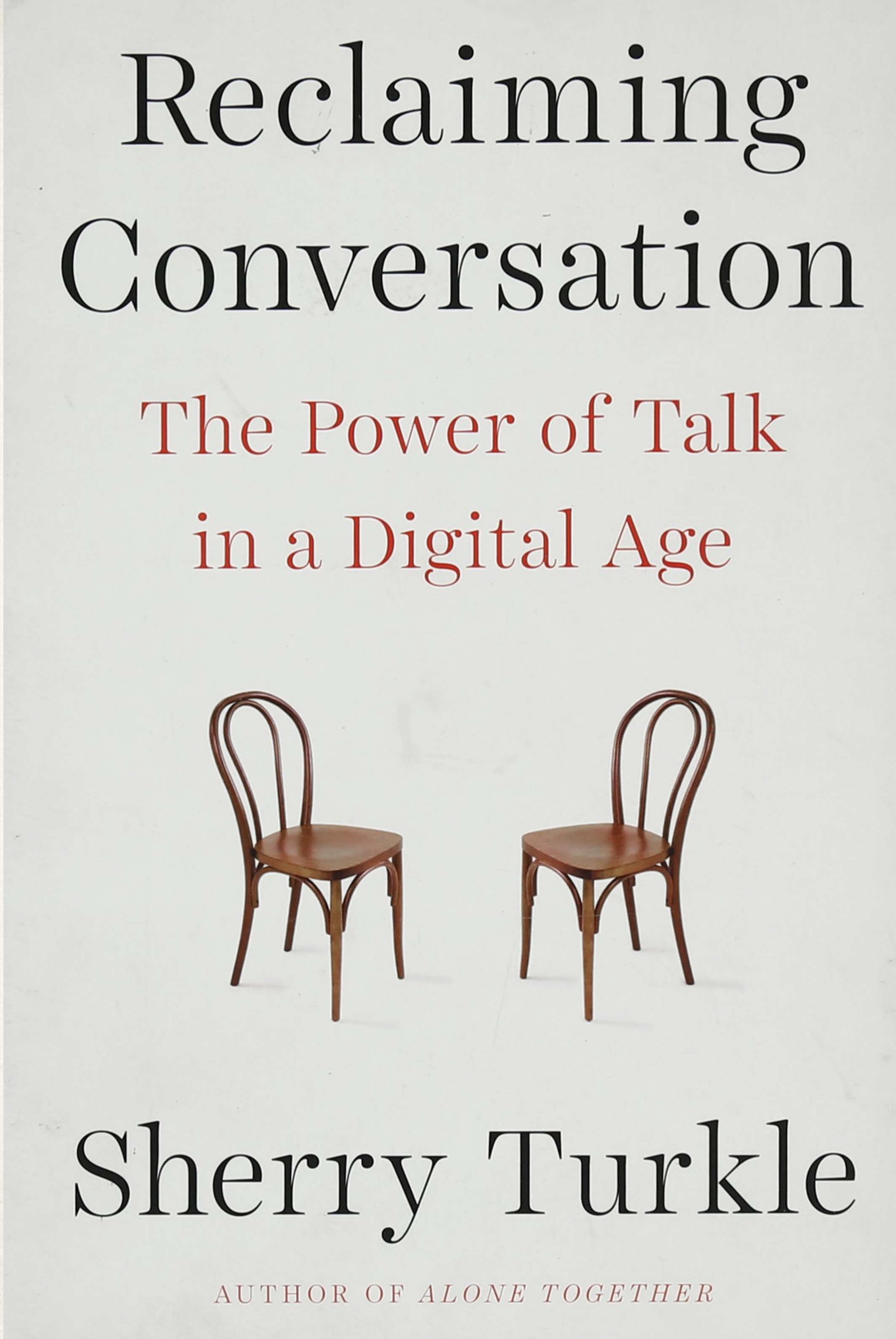
Challenges, Bias and Ethical implications

Content analysis (NLP) + (big) social media data = new responsibilities



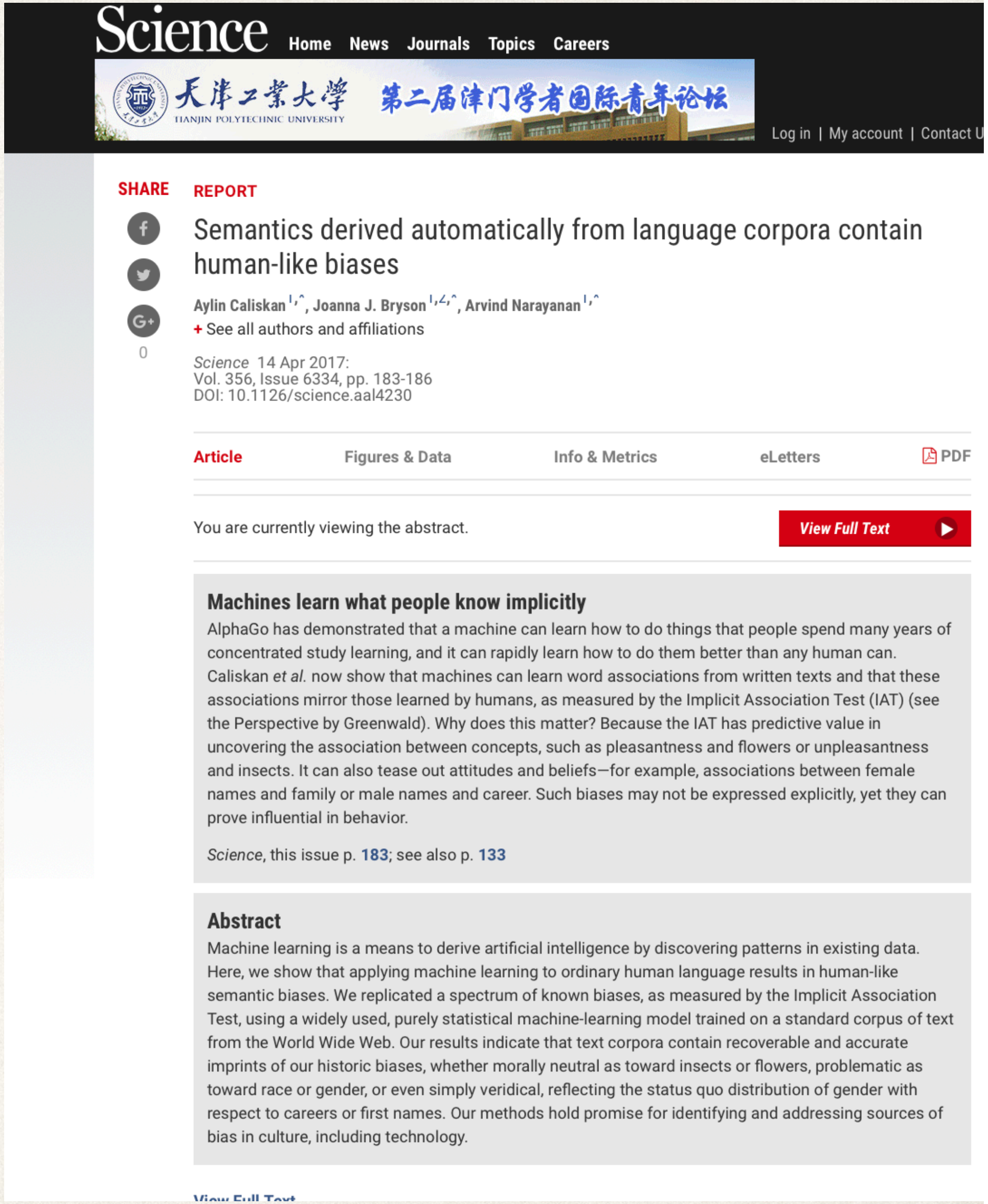
Ethical implications: Human Bias, Prejudices and Stereotypes

- Corpora of texts from social media contain recognizable footprints of the historical prejudices of the users community.
  - Users express themselves using an unfiltered style, in a communicative context that inhibits our very human capacity to be empathic - typical of interpersonal contexts (Turkle, 2017).



Ethical implications: Human Bias, Prejudices and Stereotypes

- Machine learning algorithms trained on spontaneous texts produced by a community which, even implicitly, expresses a prejudicial attitude towards people or groups of people, learn a semantics imbued with the stereotypes and the human bias of that community and of society in general (Caliskan, Bryson et al, 2017)



Science Home News Journals Topics Careers

TIANJIN POLYTECHNIC UNIVERSITY 第二届津门学者国际青年论坛

Log in | My account | Contact U

SHARE

0

REPORT

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]

+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 183-186
DOI: 10.1126/science.aal4230

Article | Figures & Data | Info & Metrics | eLetters | PDF

You are currently viewing the abstract. **View Full Text**

Machines learn what people know implicitly

AlphaGo has demonstrated that a machine can learn how to do things that people spend many years of concentrated study learning, and it can rapidly learn how to do them better than any human can. Caliskan *et al.* now show that machines can learn word associations from written texts and that these associations mirror those learned by humans, as measured by the Implicit Association Test (IAT) (see the Perspective by Greenwald). Why does this matter? Because the IAT has predictive value in uncovering the association between concepts, such as pleasantness and flowers or unpleasantness and insects. It can also tease out attitudes and beliefs—for example, associations between female names and family or male names and career. Such biases may not be expressed explicitly, yet they can prove influential in behavior.

Science, this issue p. 183; see also p. 133

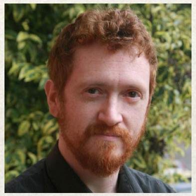
Abstract

Machine learning is a means to derive artificial intelligence by discovering patterns in existing data. Here, we show that applying machine learning to ordinary human language results in human-like semantic biases. We replicated a spectrum of known biases, as measured by the Implicit Association Test, using a widely used, purely statistical machine-learning model trained on a standard corpus of text from the World Wide Web. Our results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral as toward insects or flowers, problematic as toward race or gender, or even simply veridical, reflecting the status quo distribution of gender with respect to careers or first names. Our methods hold promise for identifying and addressing sources of bias in culture, including technology.

View Full Text

Bias in AI e NLP? We need diversity!

- Bias in developing resources and annotated corpora to be used as training and testing data
    - Definition of the phenomena we want to model (e.g. hate speech)
    - Selection of training data (source, authors,…)
    - Biases of the annotators
        - We need to deal with human diversity!
- Machine learning with a point of view?
- Perspectivist data Manifesto: https://pdai.info/



Ethical implications

- Development and growth of a deeper awareness about the ethical aspects related to the application of language and communication technologies to the analysis of large amounts of user-generated contents
    - Especially when sentiment analysis is applied & combined with author profiling
- Researchers and technology companies must be well-equipped with awareness & sense of responsibility
    - Machine learning models tend to incorporate the bias present in human behavior.
- Not only risks but also opportunities:
    - Bias and stereotypes can be observed and analysed “in the wild” in spontaneous dialogues among social media users: hate speech detection, deception detection and so on can be precious to provide an ethical contribution in terms of knowledge that can be used to counteract negative trends.

Research Ethic Statements And Data Sheets

EMNLP 2020 Plenary Panel Discussion:

Publishing in the Era of Responsible AI: How Can we be Proactive? Considerations and Implications.

Emily M. Bender, Rosie Campbell, Allan Dafoe, Pascale Fung, Meg Mitchell, Saif M. Mohammad

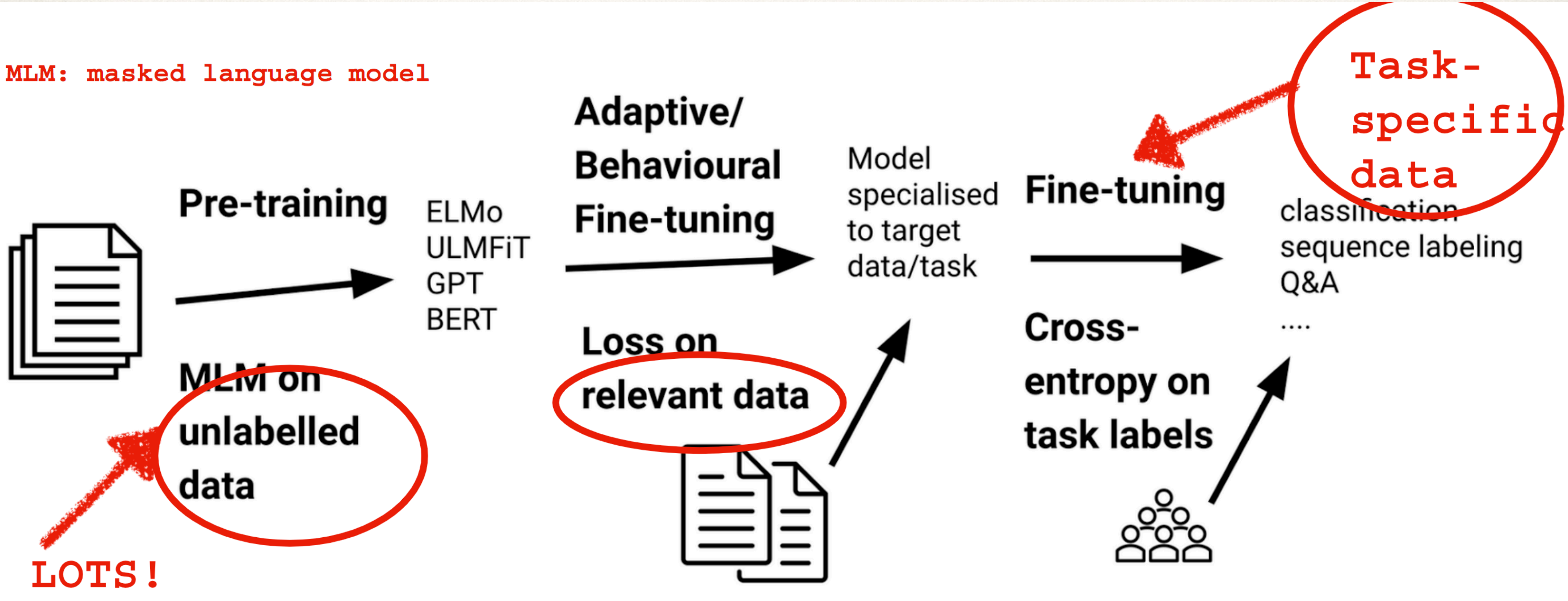
What is a Research Ethics Statement and Why does it Matter?

Saif M. Mohammad

Senior Research Scientist, National Research Council Canada

Saif.Mohammad@nrc-cnrc.gc.ca @SaifMMohammad

Base Models + Further Training + Fine-Tuning



Revolutionised everything, from generation to MT, and lots of other tasks BUT, bias in data

https://ruder.io/recent-advances-lm-fine-tuning/

LLMs scenario

- Neutral? Not really :)
- Warning: Data-driven (neural) NLP systems
- Diversity (authors…)

Smaller (curated) resources

Large web-mined sets

Increasing work establishing that data is never 'raw', or abstract but rather is shaped through the practices of collecting, curating and sensemaking, and thus is inherently sociopolitical in nature.