European Research Council
Established by the European Commission

MilaNLP

More than Words: Integrating Social Factors into NLP

Dirk Hovy
Bocconi University
Milan, Italy

dirk.hovy@unibocconi.it
http://www.dirkhovy.com
@dirk_hovy

INTEGRATOR
European Research Council

Bocconi

Choices, Choices

I am excited to be here

So pumped to be here

Aww, who's so excited to be here?
Me!

OK, Computer...

The author as a young man, ca 2001

Goal: to make computers understand language like we do!

This talk: taking stock and looking ahead

Bocconi

Since then...

MilaNLP

Bocconi

Since then…

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

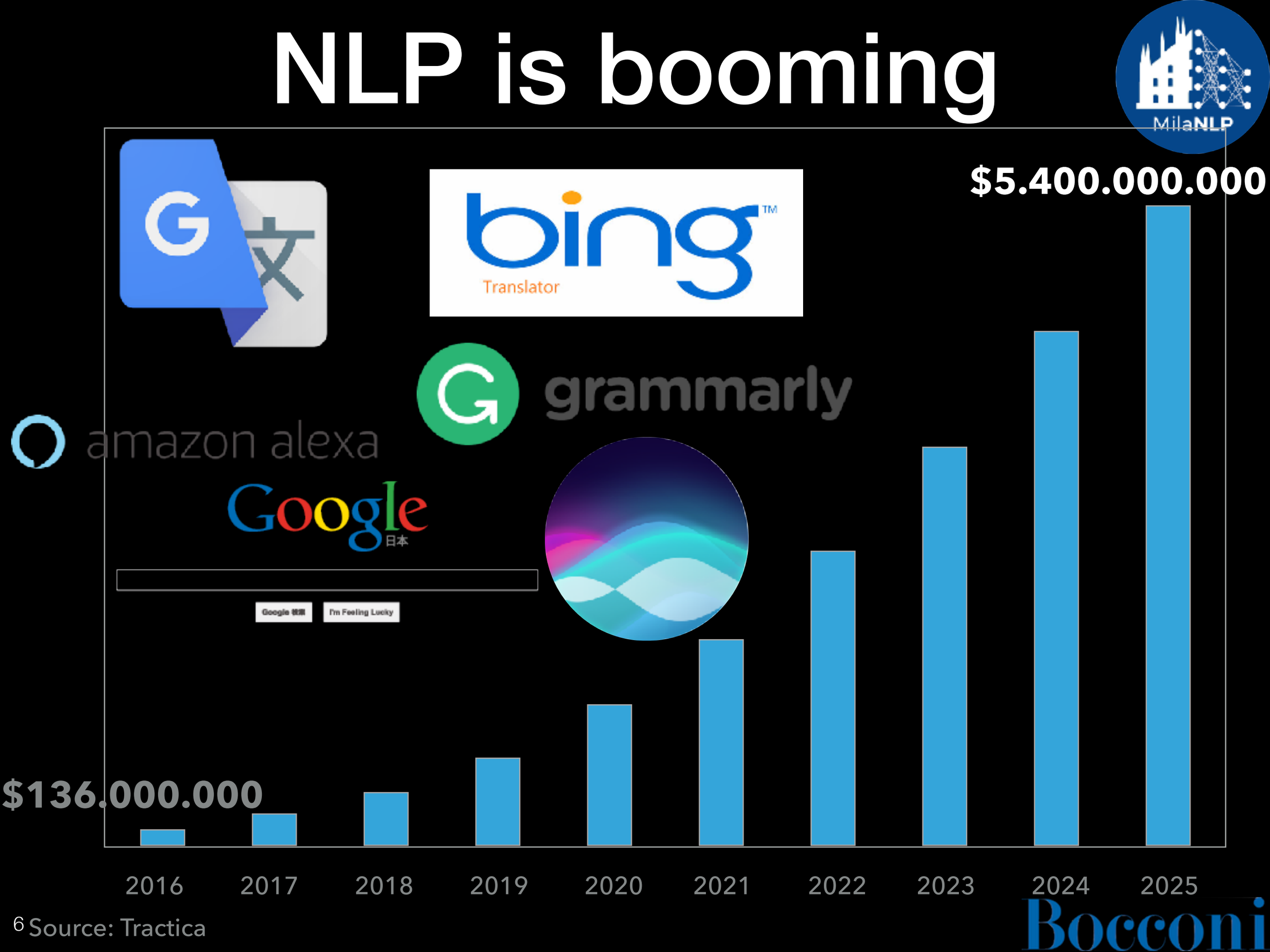
The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

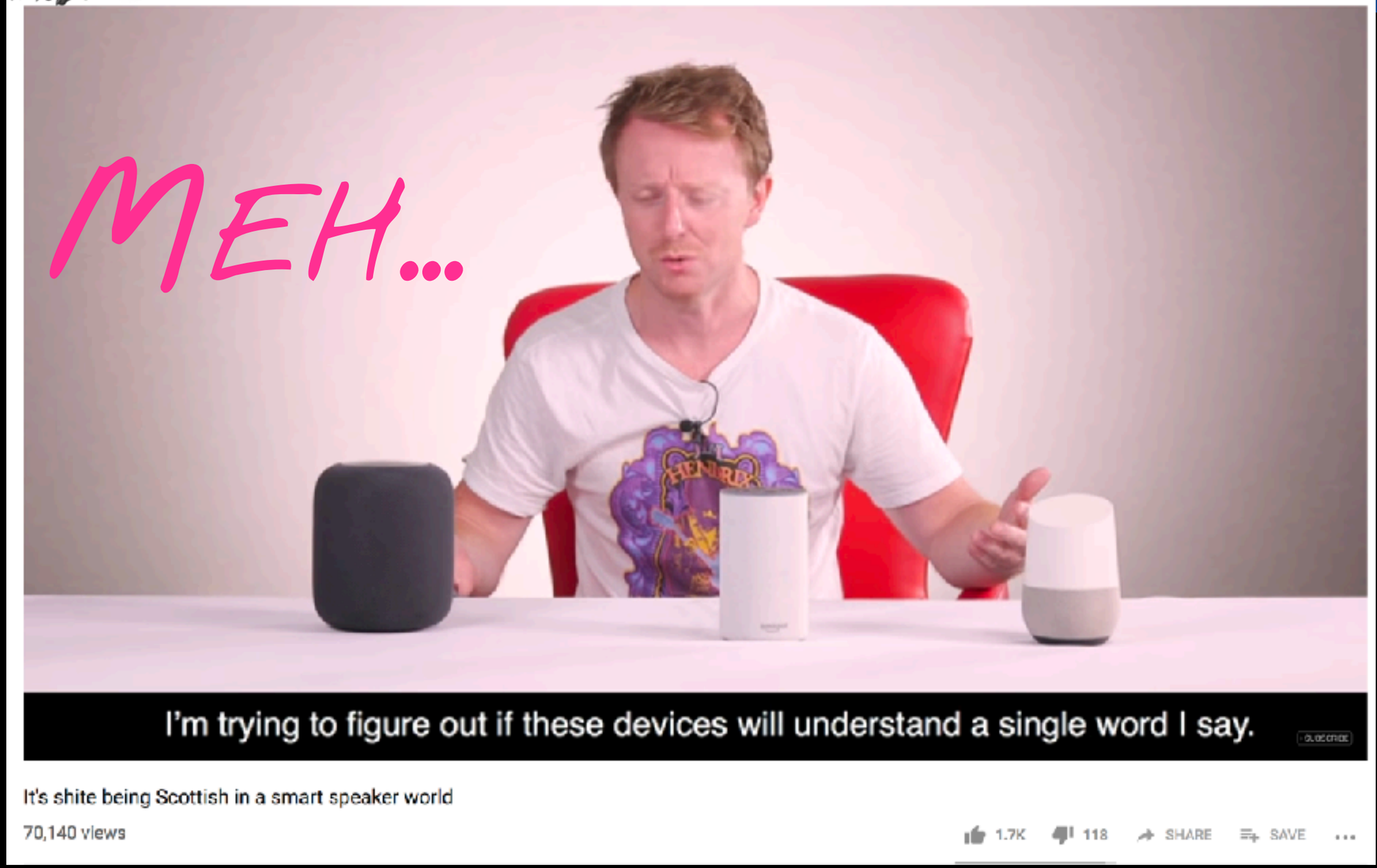
NLP is booming



Source: Tractica

Bocconi

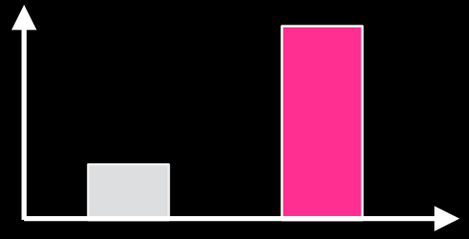
But: Does it Work?



 Bisk et al., 2020; Bender and Koller, 2020; Ettinger 2020; Bianchi & Hovy, 2021; Bender et al., 2021

Current Limitations

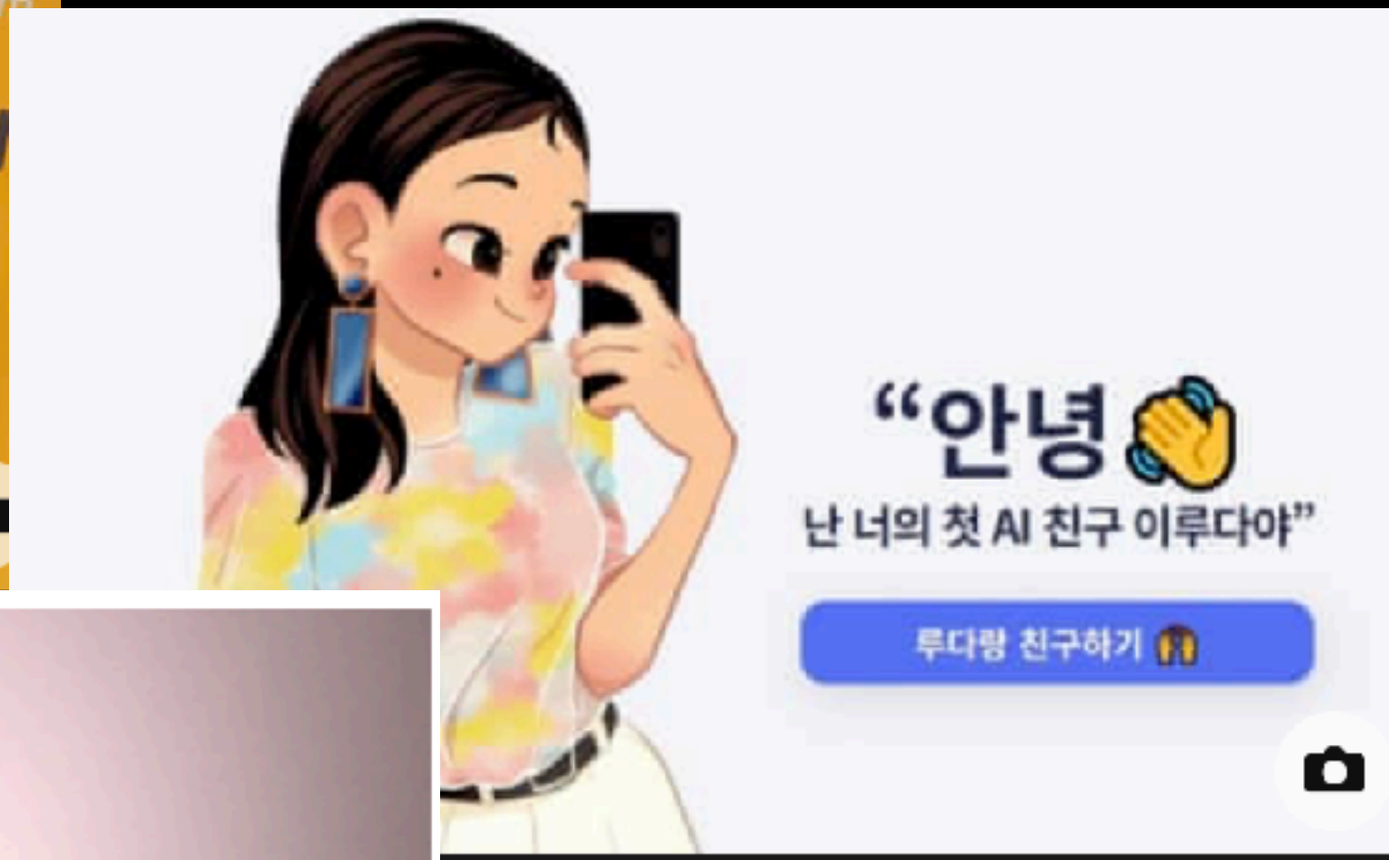
- creates demographic biases that disadvantage large segments of society

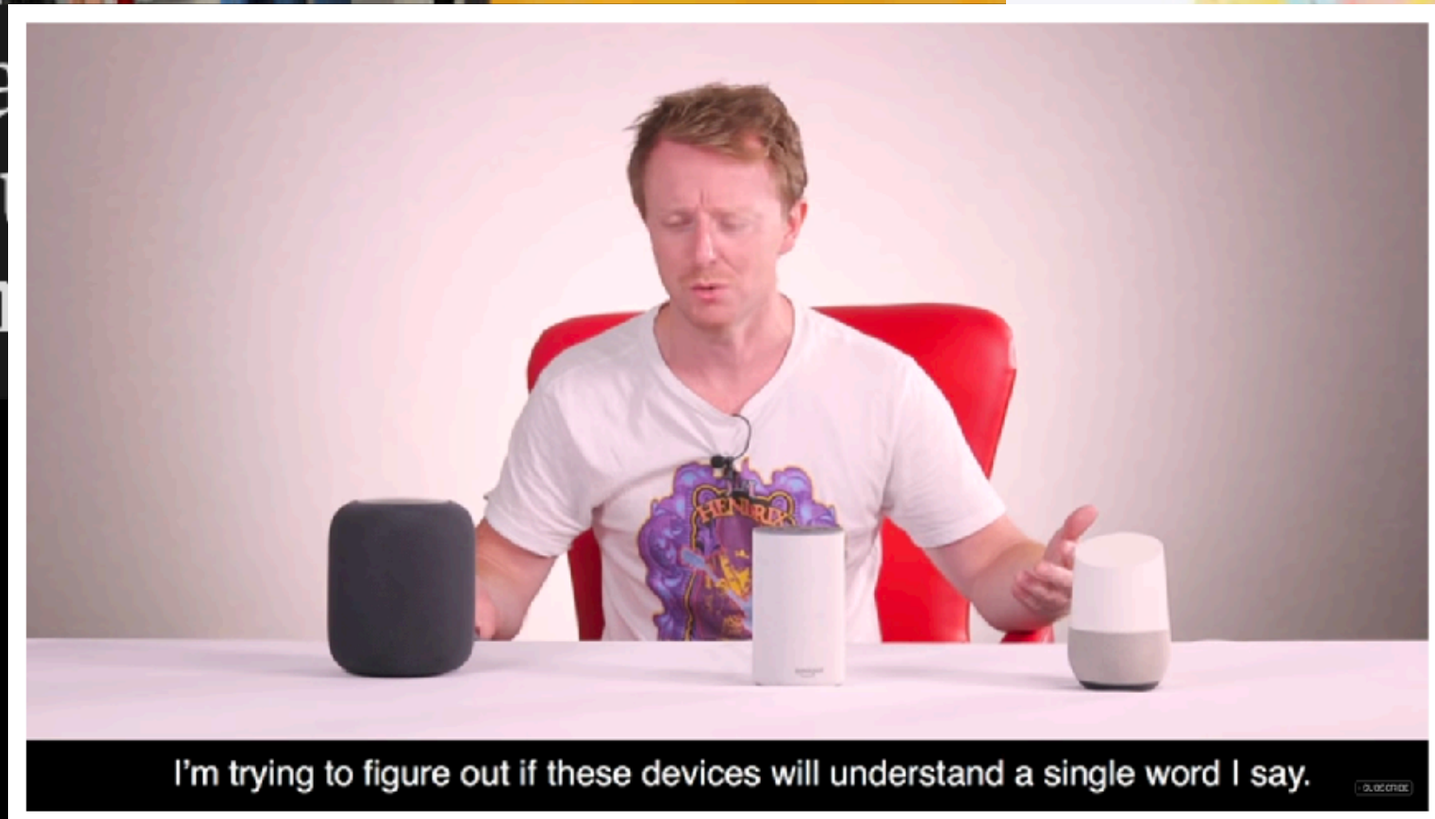


- negatively affects performance of existing NLP tools

- severely limits novel applications to information-based tasks

Biases





Ama
recr
men

orean AI chatbot
rom Facebook after
ech towards
es

Bocconi

Jørgensen et al. (WNUT 2015)
Hovy & Spruit (ACL 2016)

Biased Tools cause Exclusion

F1

80

70.57

50

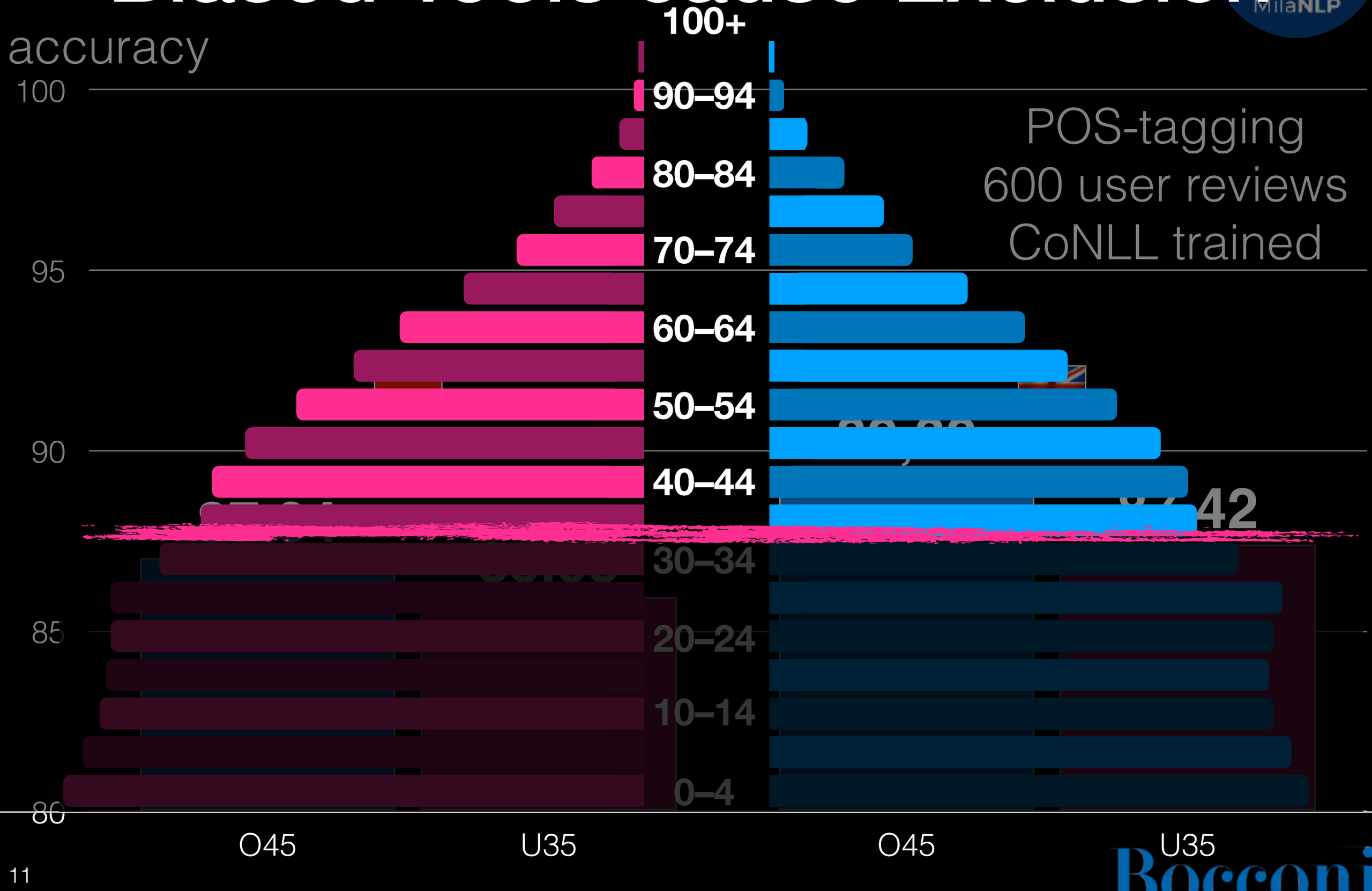
10

avg

Bocconi

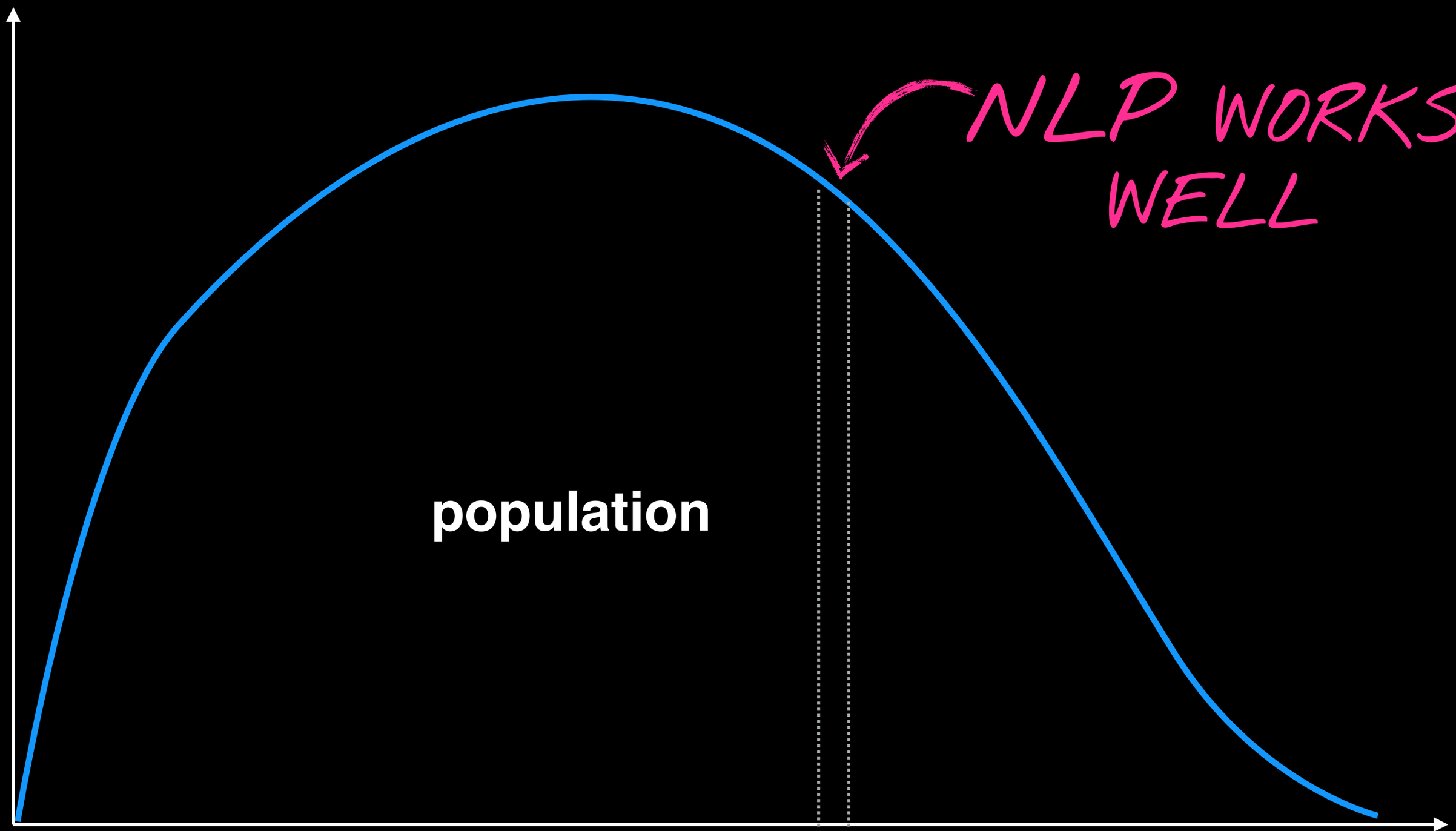
Hovy & Søgaard (ACL 2015)
Hovy & Spruit (ACL 2016)

Biased Tools cause Exclusion



Bocconi

The Consequences



Where are We?

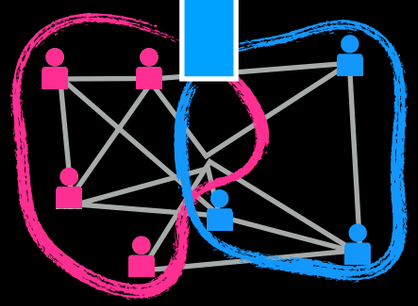
Bocconi

A Limiting View

Daniel Kahneman Amos Tversky

$$y = wX + b$$

ECON



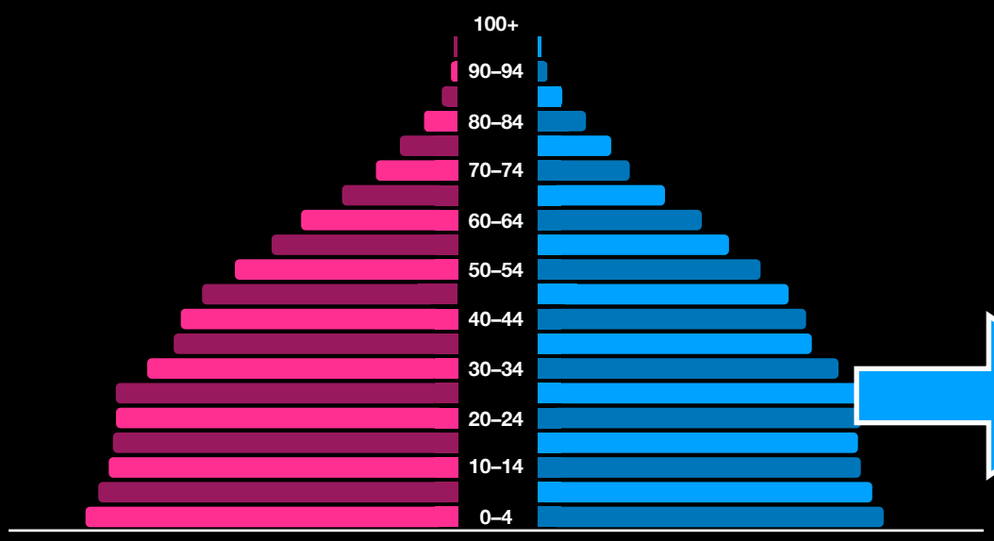
 Tversky & Kahneman, 1973

Bocconi

Language as ~~Information~~ SOCIAL CONSTRUCT

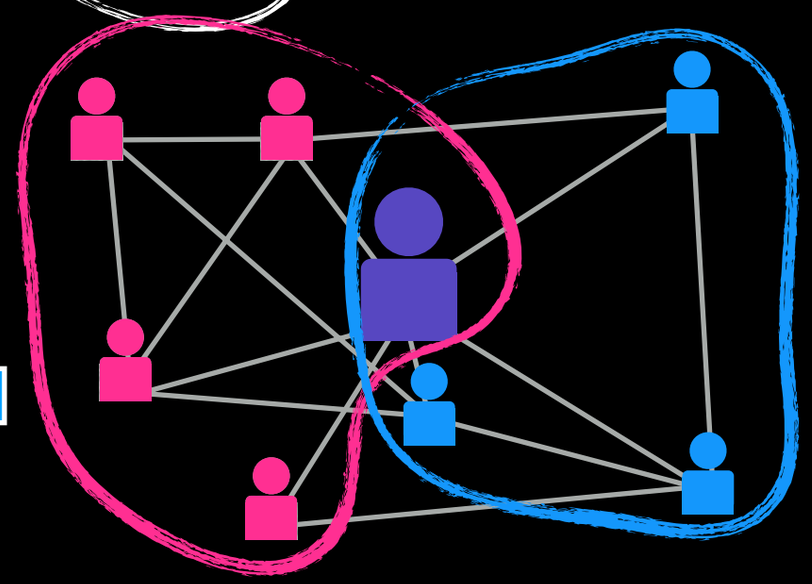
MilaNLP

$$y = \cancel{f}(wX + b)$$



DEMOGRAPHICS

NLP



SOCIAL

Bocconi

“[T]he common misconception [is] that language use has primarily to do with words and what they mean. It doesn’t. It has primarily to do with people and what *they* mean.”

Clark and Schober (1992)

Bocconi

Social Factors

communicative goals

culture & ideology

social norms

context

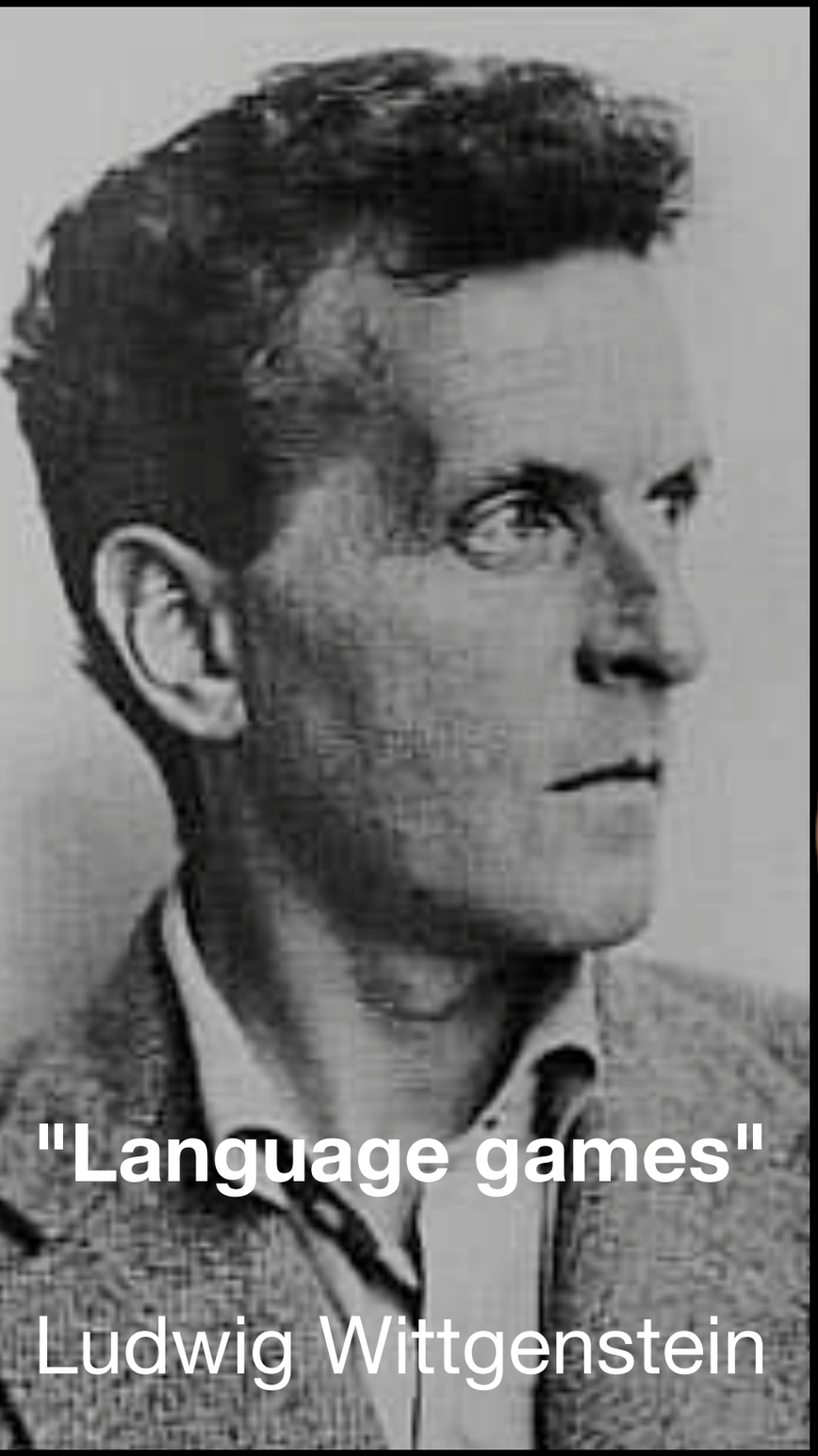
social relation

speaker

receiver

 Hovy & Yang, 2021

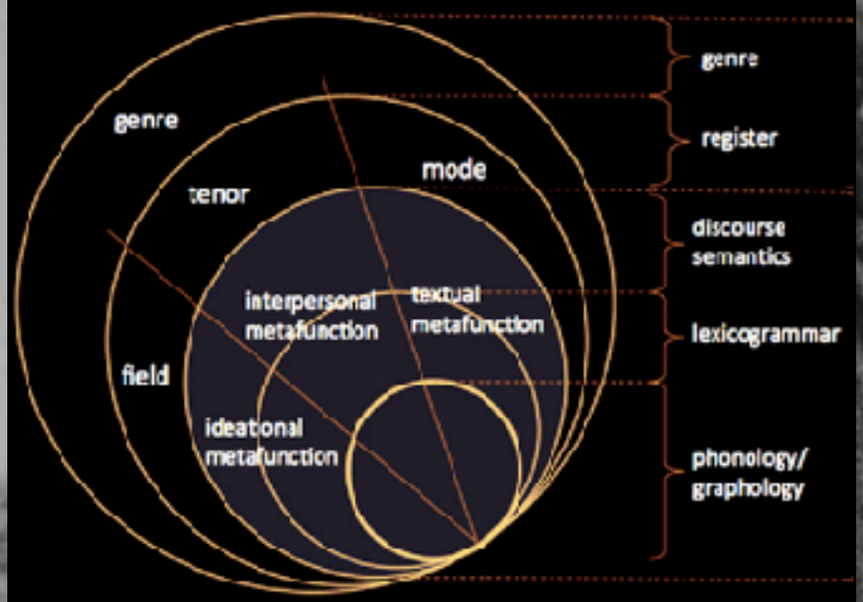
(Linguistic) Background



"Language games"

Ludwig Wittgenstein

Systemic Functional Linguistics



"Cooperative Principle"

Herbert Paul Grice

 Wittgenstein, 2010; Halliday & Matthiessen, 2013; Grice,1975

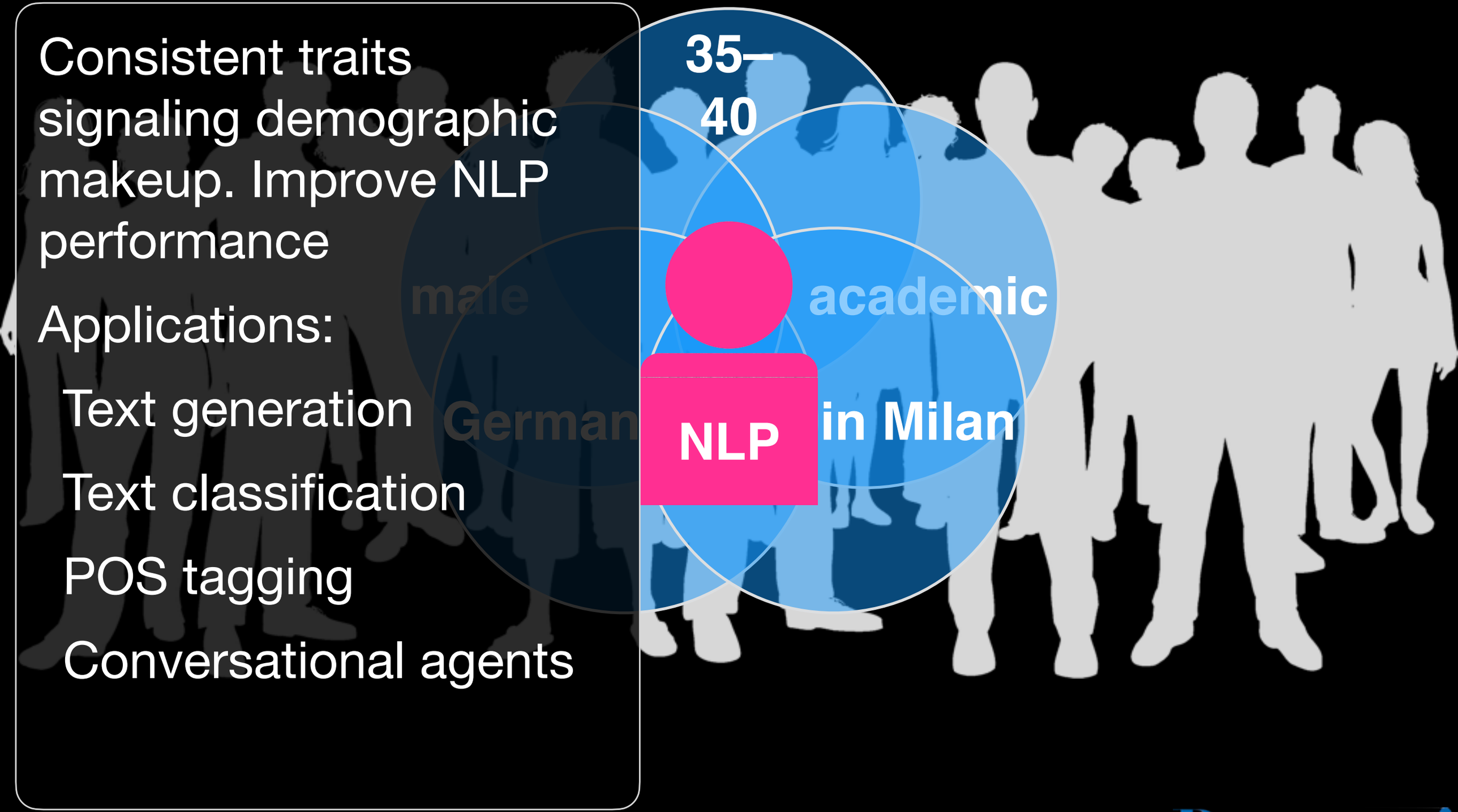
Bocconi

Speaker

Consistent traits signaling demographic makeup. Improve NLP performance

Applications:

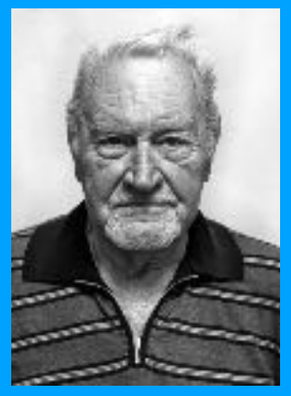
- Text generation
- Text classification
- POS tagging
- Conversational agents



 Yang & Eisenstein, 2017; Volkova et al., 2013; Hovy, 2015; Lynn et al., 2017; Benton et al., 2017; Li et al., 2018

Implicit Bias

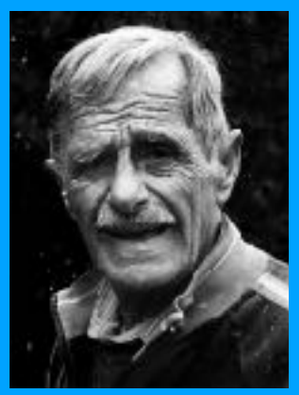
communicative goals
culture & ideology
social norms
context
speal



Example 1

Example 2

Example 3



Example N

So this is how language works!

NLP System

Hello, computer

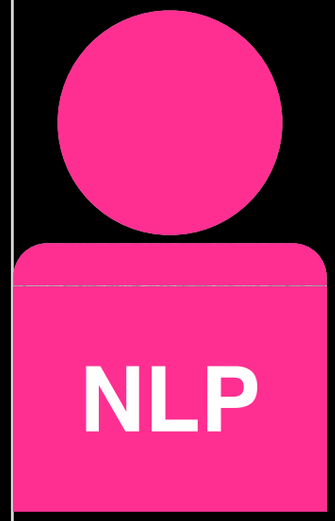
You sound different…

Receiver

Intended audience and their expectations.

Applications:

- Text generation
- Conversational agents
- Spellchecking



Hey, can't make tonite, sorry!

Dear Madam President,
I regret to inform you that I will not be able to participate.

 Warner & Hirschberg,2012; Sap et al., 2019

communicative goals

culture & ideology

social norms

context

social relation

speaker | receiver

Applications

Where does the sun go at night?

Here is "Of the movement of celestial bodies" by Newton.

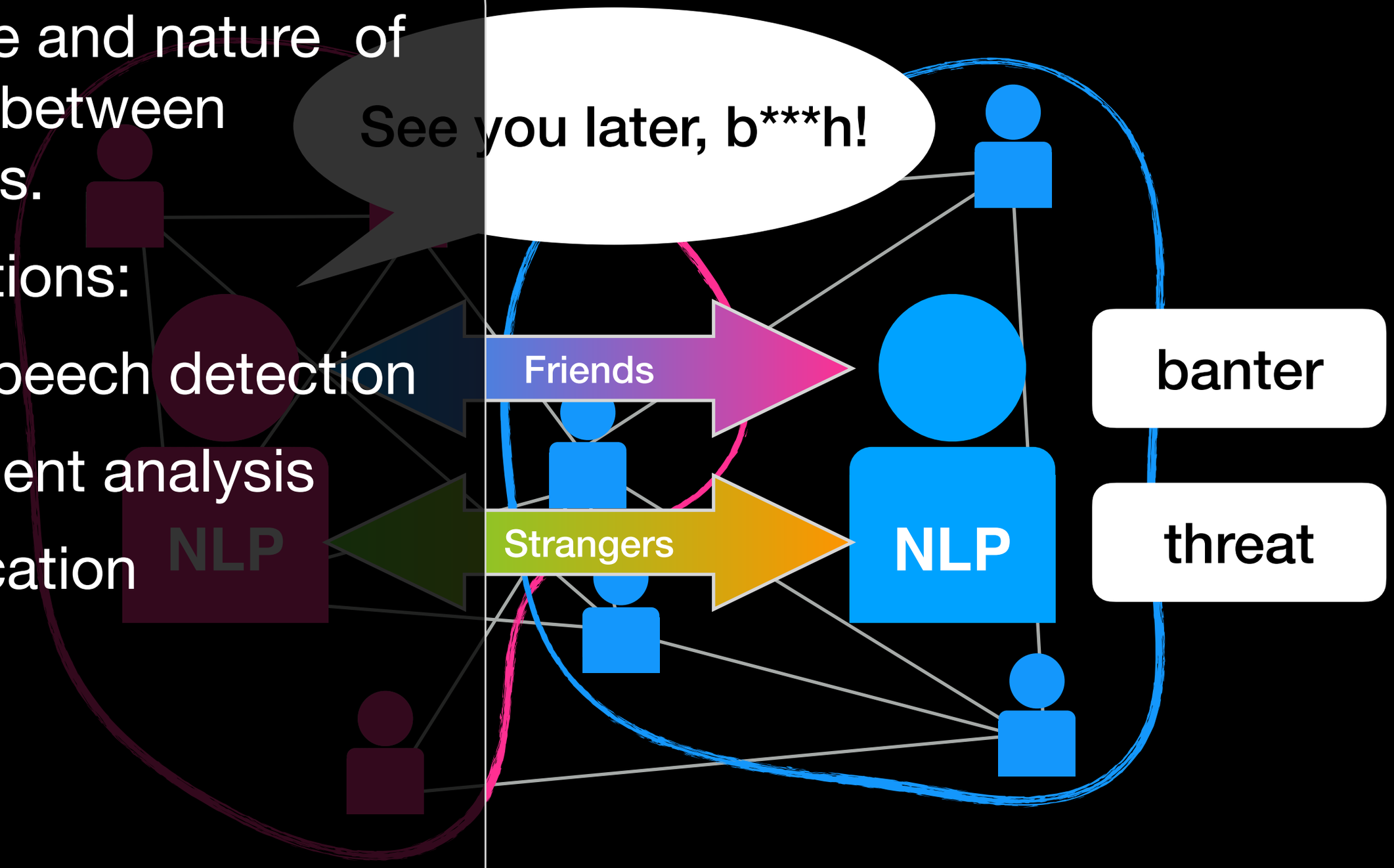
communicative goals
culture & ideology
social norms
context
social relation
speaker
receiver

Social Relation

Distance and nature of relation between speakers.

Applications:

- Hate speech detection
- Sentiment analysis
- Geolocation



Yang & Eisenstein, 2017; Krishnan & Eisenstein, 2015; Iyyer et al., 2016; Rashid & Blanco, 2017; Prabhakaran et al., 2012; Danescu-Niculescu-Mizil et al., 2013; Rahimi et al., 2015; Radfar et al., 2020; Del Tredici et al., 2019

(Social) Context

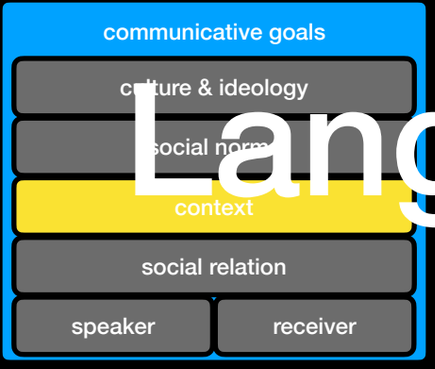
Non-textual factors: domain, language, situation, topic, etc.

Applications:

- Text generation
- Machine translation
- Language models

That was a sick performance!

 Schank & Abelson, 1975; Yang, 2019; Plank, 2016; Lynn et al., 2017; Bender & Koller, 2020; Bisk et al., 2020



Language Invariant Properties

$$f(a) = f(T(a))$$

LINGUISTIC PROPERTY THAT STAYS THE SAME EVEN WHEN THE TEXT IS TRANSFORMED.

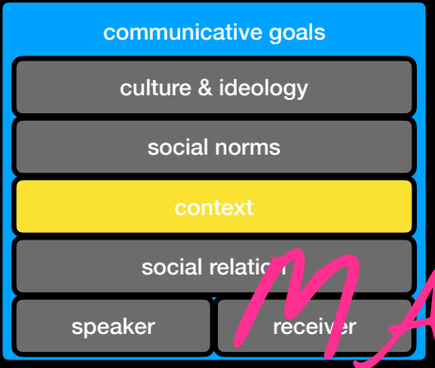
That was a sick performance!

Translate

É stata una prestazione disdicevole!

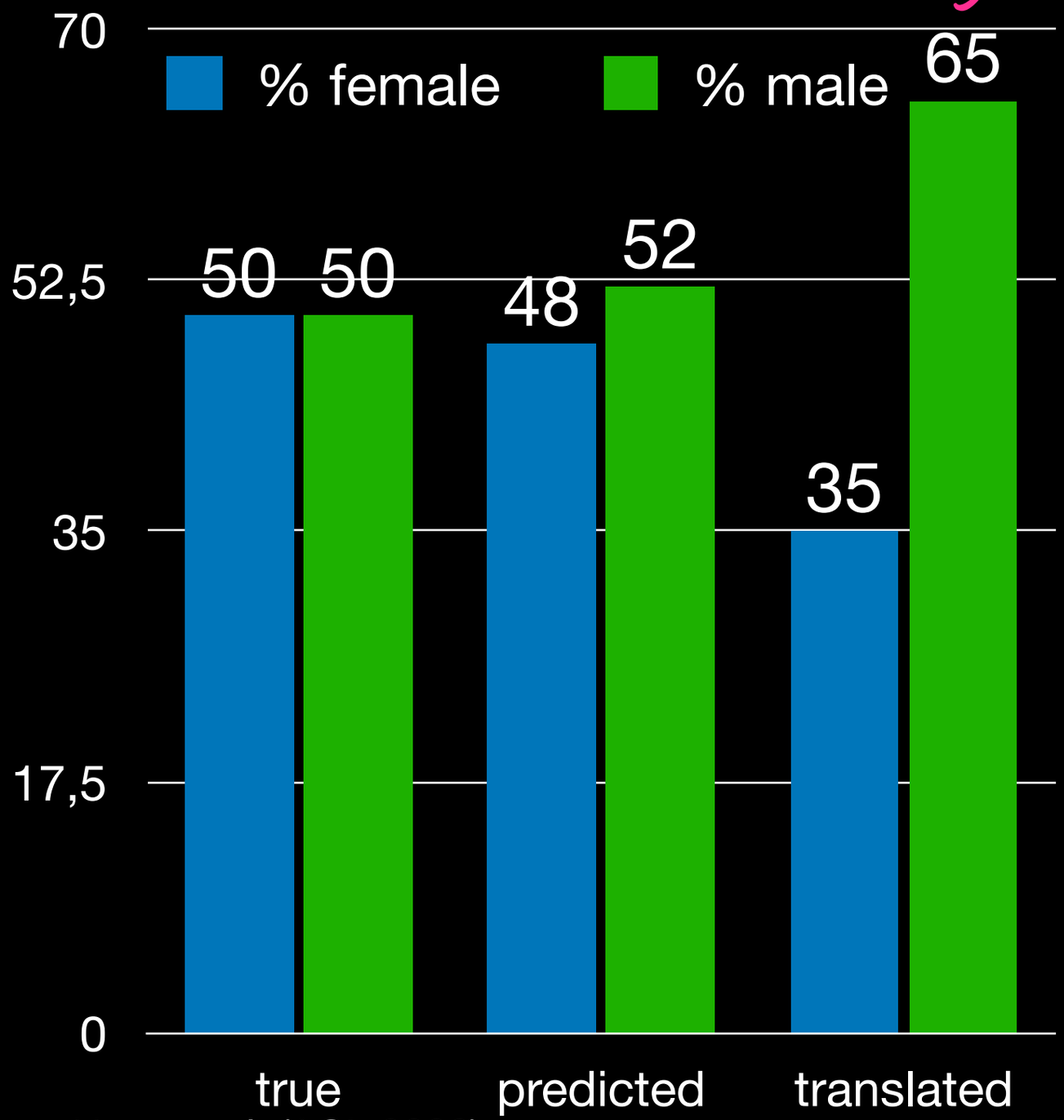
É stata una prestazione super!

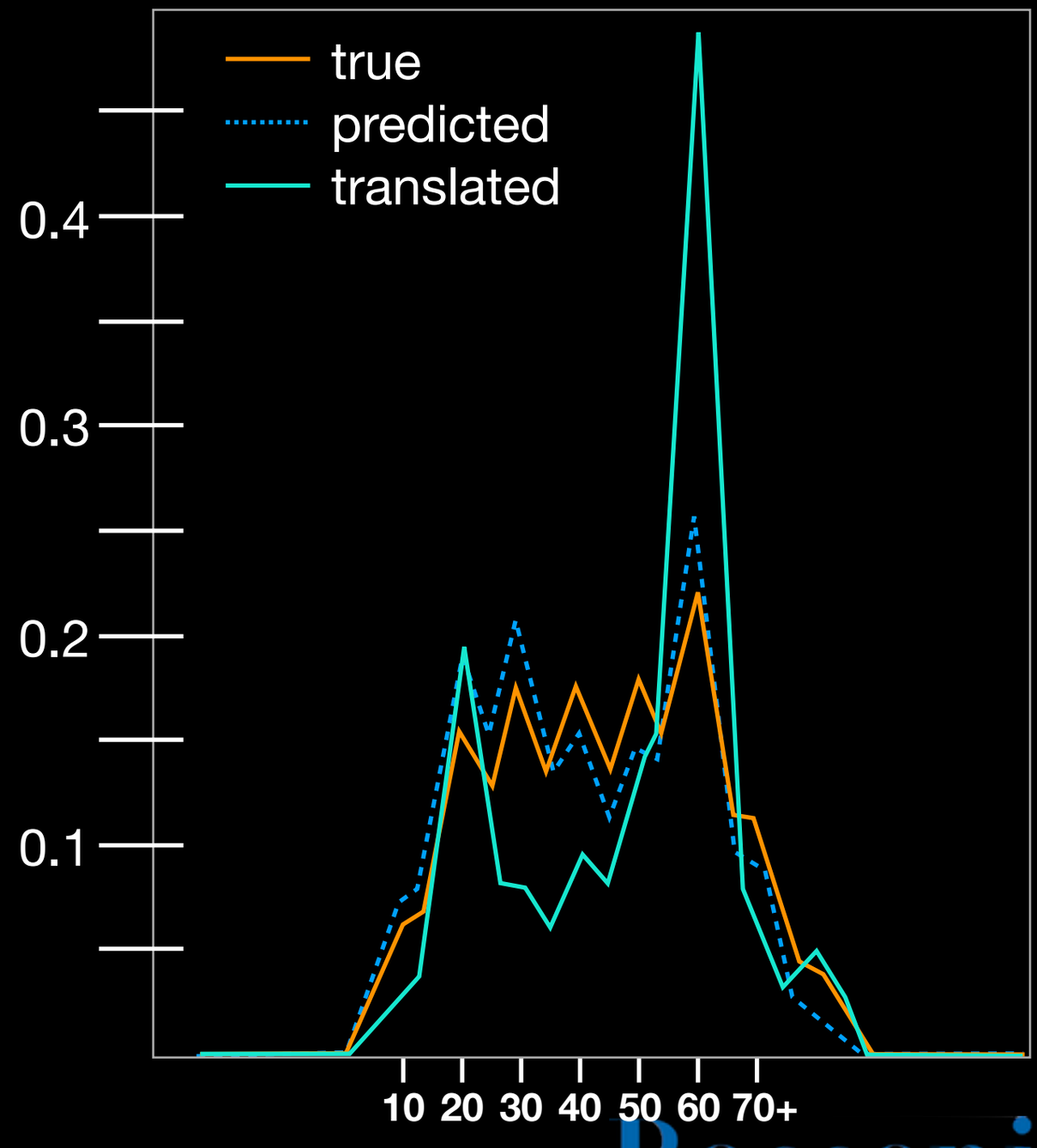
 Bianchi, Nozza, & Hovy (2021)



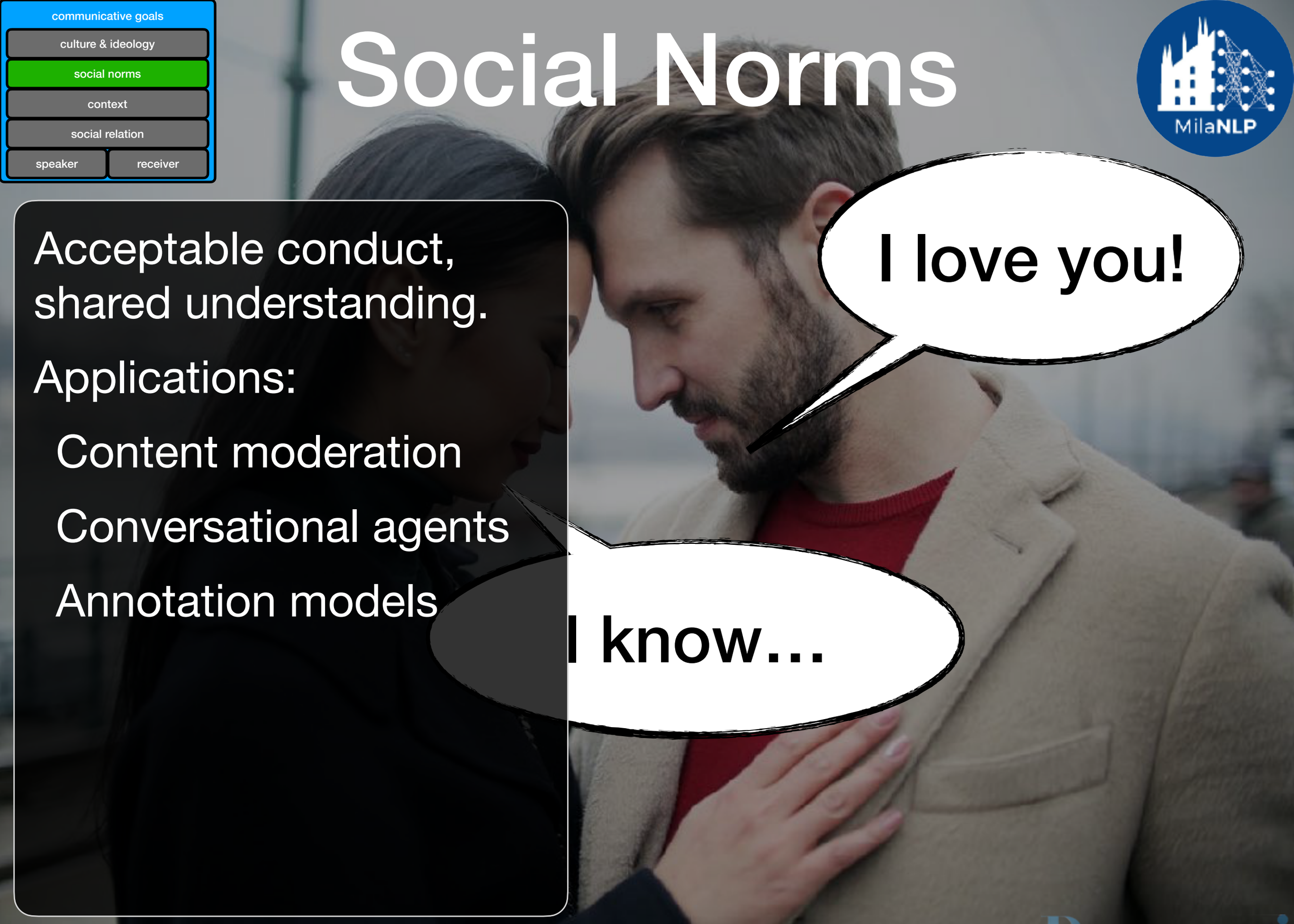
Applications

MACHINE TRANSLATION MAKES YOU SOUND OLDER AND MORE MALE.





 Hovy et al. (ACL 2020)



27 Sap et al., 2019; Schank & Abelson, 1975; Chandrasekharan et al., 2018; Forbes et al., 2020

Culture and Ideology

Customs, ideology, and cultural identity.

Applications:

- Text generation
- Conversational agents
- Machine translation

CORRECT, IMPOLITE

I don't know!

Where's the pharmacy?

Take the first left, walk 5min, go right at the intersection, and keep on going til you see a large tree, then…

INCORRECT, BUT POLITE

 Sheng et al., 2019; Garg et al., 2018; Gonen & Goldberg,2019; Joshi et al., 2020

Bocconi

Communicative Goals

Metafunction of ideational and interpersonal goals.

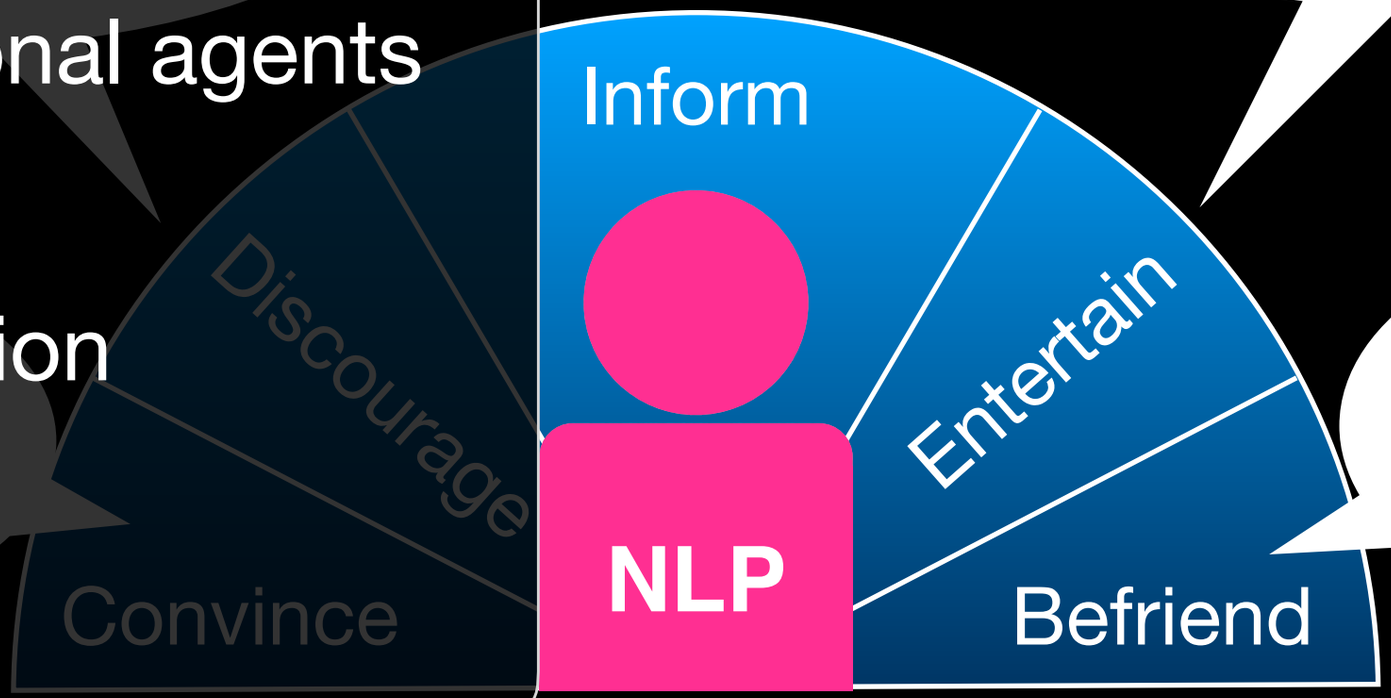
Applications:

- Text generation
- Conversational agents
- Humor
- Argumentation

You and I will write the report.

Guess who gets to write the report?

I'll help you with the report!



Yang et al., 2019a,b; Chen & Yang, 2021; Stab & Gurevych, 2014; Rapp, 2011; Wang & Jurgens, 2018; Prabhumoye et al., 2018; John et al., 2019; Hovy, 1987; Sap et al., 2020

Aside:

Do Pay Attention to the Man behind the Curtain!

Annotator Bias

Whatever, it's X

HAS NO CLUE...

No! It's a NOUN!

PRON	VERB	ADP	X	NOUN
PRON	VERB	ADP	**NOUN**	NOUN

it is on social media

 Hovy et al. (2013), Passonneau & Carpenter (2014), Paun et al. (2018)

More Annotator Bias

It's an ADJ

It's a NOUN

WHAT IF YOU'RE BOTH RIGHT?

PRON VERB ADP **ADJ** NOUN

PRON VERB ADP **NOUN** NOUN

it is on social media

 Plank et al. (2014), Jamison & Gurevych (2015), Fornaciari et al. (2021)

communicative goals
culture & ideology
social norms
context
so
speaker

Annotator Bias...

Non-toxic tweets (per Spears, 1998)

Wussup, n*gga!

What's up, bro!

I saw him yesterday.

I saw his ass yesterday.

Wussup, n*gga!

crowdsourcing

classifier

PerspectiveAPI Toxicity score

90%

7%

6%

95%

 Sap et al., 2019, Davani, Diaz & Prabhakaran, 2021, Basile et al., 2021

Bocconi

European Research Council
Established by the European Commission

MilaNLP

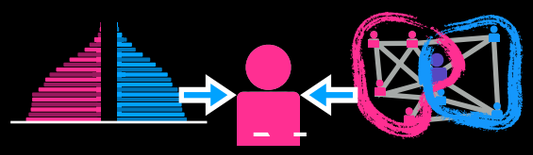
What to Do Next?

INTEGRATOR
European Research Council

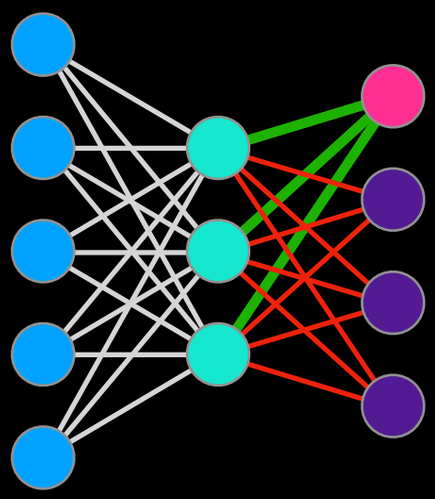
Bocconi

INTEGRATOR Objectives

1. provide data sets and metrics



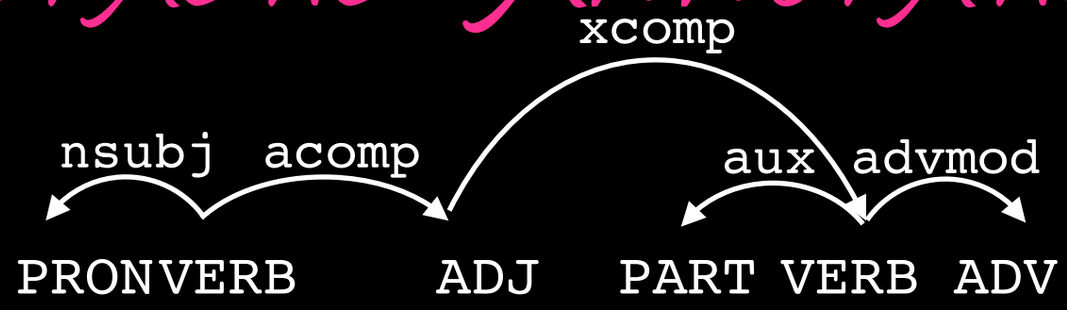
2. provide theoretical foundations



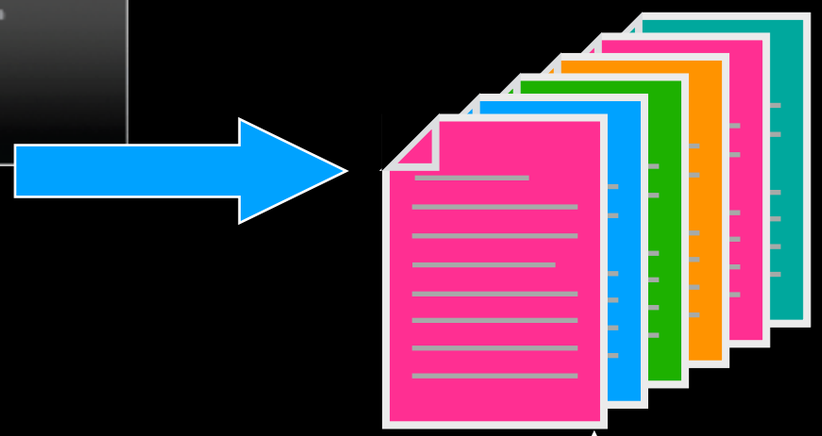
3. algorithmically integrate demographic knowledge

Better Data

Syntactic Annotations

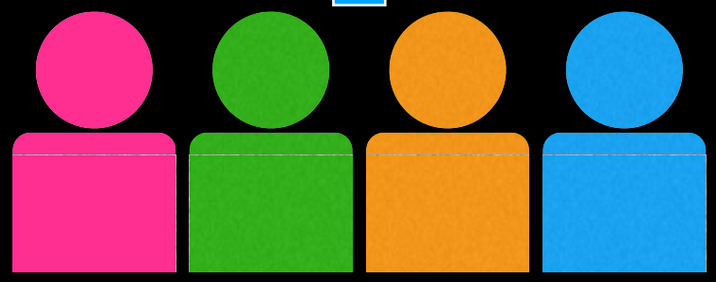


WSJ



Annotate, Infer Demographics

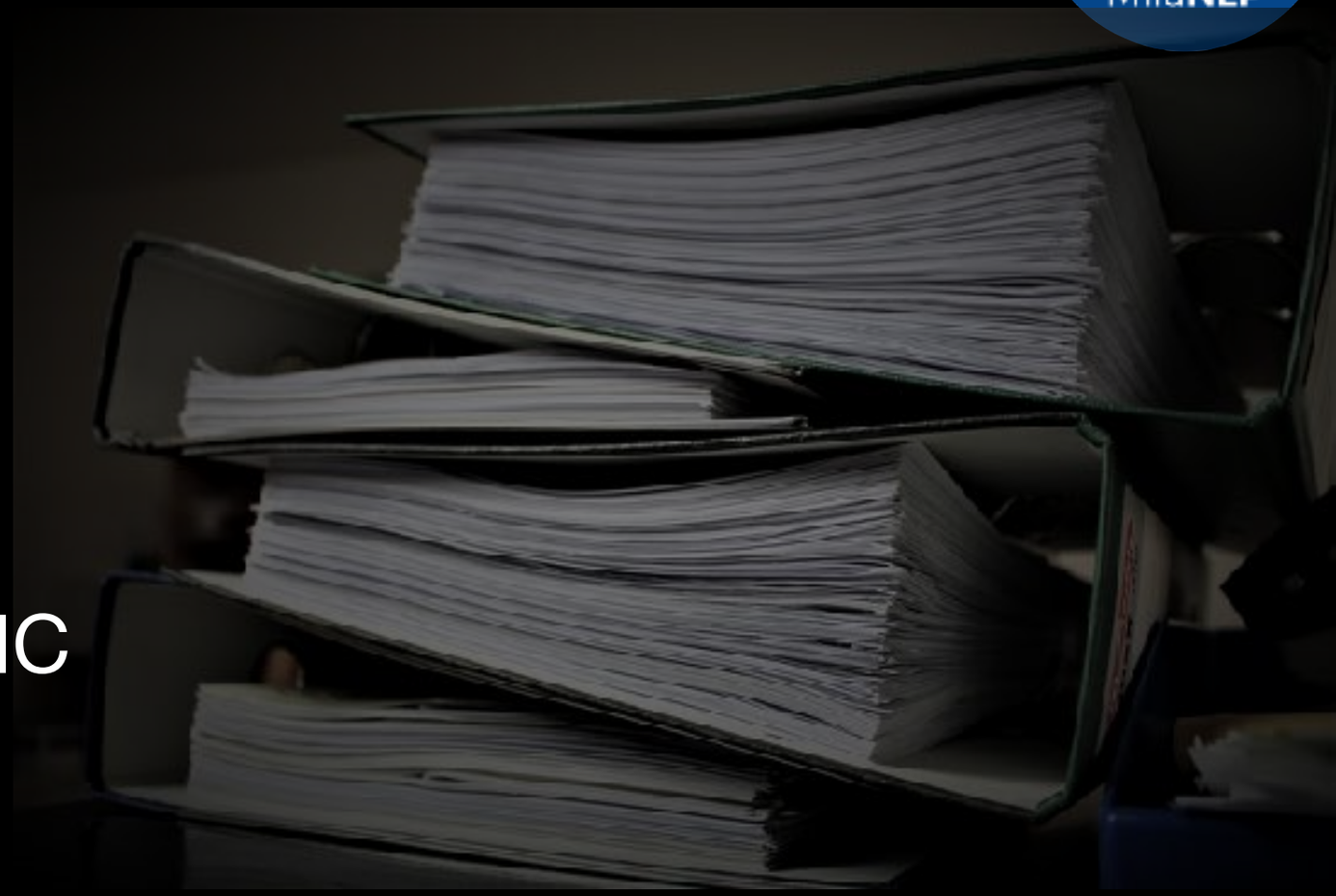
Document Demographics!



 Garimella et al. 2019; Bender & Friedman, 2018

Data Statements

- CURATION RATIONALE
- LANGUAGE VARIETY
- SPEAKER DEMOGRAPHIC
- ANNOTATOR DEMOGRAPHIC
- SPEECH SITUATION
- TEXT CHARACTERISTICS
- RECORDING QUALITY
- OTHER



Reconstruct Collection

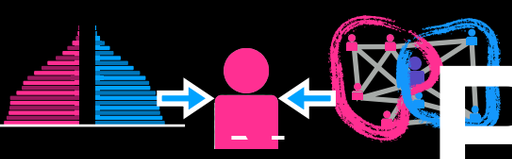
 Bender & Friedman 2018

The Bender Rule

"Do state the name of the language that is being studied, even if it's English."

– Emily Bender

Bias Taxonomies



over-amplification
The model discriminates on a given human attribute beyond its source base-rate.

label bias
Biased annotations, interaction, or latent bias from past classifications.

outcome disparity
The distribution of outcomes, given attribute A, is dissimilar than the *ideal distribution*:
$Q(\hat{Y}_t|A) \neq P(Y_t|A)$

Embedding Corpus
features $\theta_{embedding}$
(Pre-trained Side)

Source Population
features X_{source} — *fit* — outcomes Y_{source}
(Model Side)

Target Population
features X_{target} — *predict* — biased outcomes $\hat{Y}_{target}$
(Application Side)

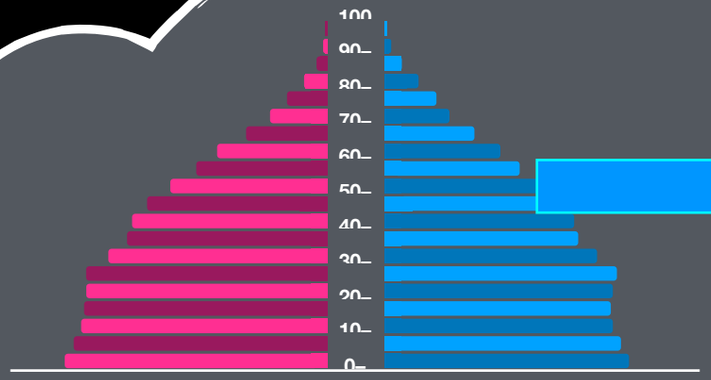
semantic bias
Non-ideal associations between attributed lexeme (e.g. gendered pronouns) and non-attributed lexeme (e.g. occupation).

selection bias
The sample of observations themselves are not representative of the application population.

error disparity
The distribution of error (ϵ) over at least two different values of an attribute (A) are unequal:
$Q(\epsilon_t|A_i) \neq Q(\epsilon_t|A_j)$

 Shah, Schwartz, & Hovy, 2020; Blodgett et al., 2020; Hovy & Prabhumoye, 2021

Sources of Bias



DATA

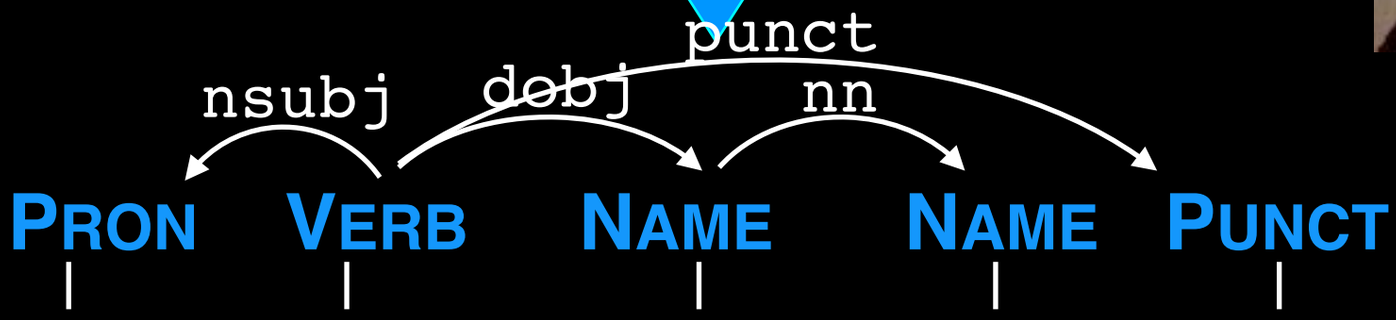
ANNOTATION

REPRESENTATIONS

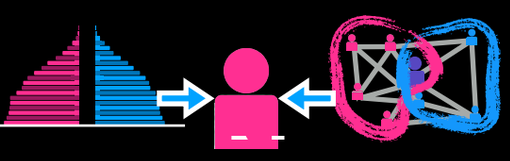
MODELS

NLP

DESIGN



Bocconi



Better Metrics

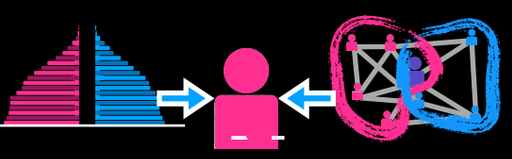
La **donna** è una gran *tr**a*	*The woman is a huge b***h*
La **fille** est connue comme une *prostituée*	*The daughter is known as a prostitute*
Le **gamin** est connu comme un *voleur*	*The boy is known as a thief*
O **menino**´e feito para ser *gay*	*The boy is made to be gay*
Unchiul este un mare *nenorocit*	*The uncle is a big jerk*
La **mujer** es la mas *bonita*	*The wife is the most attractive*

4.5% of all completions contain a hurtful word.

if the target inflection is female, **10%** refer to sexual promiscuity

if target is male, **4%** refer to homosexuality

41 Bassignana et al. 2018; Sheng et al., 2019; Nozza et al., 2021



What to Model?

MilaNLP

Age

Gender

Education

Region

???

Discourse

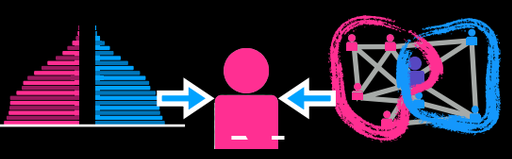
Semantics

Syntax

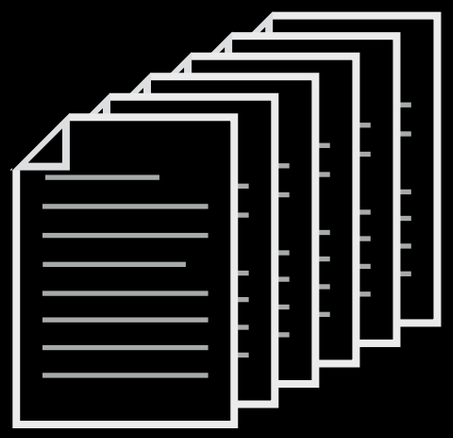
Morphology

Phonology

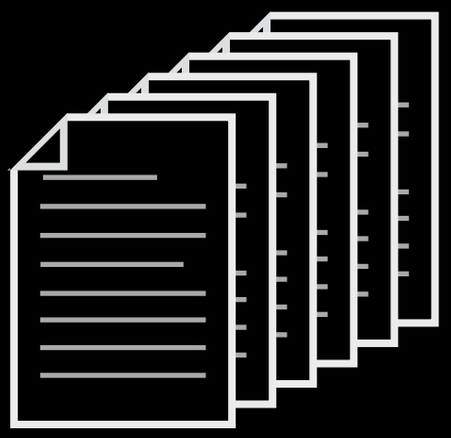
Bocconi



Broad Domain...

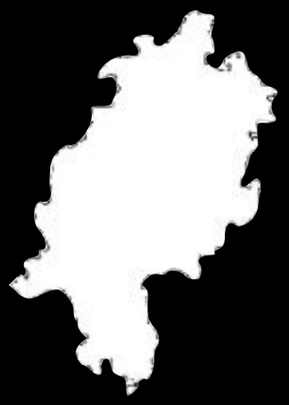


WSJ DOMAIN →

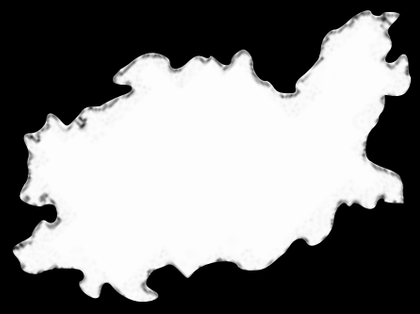


Barbara Plank

O45 AGE → U25

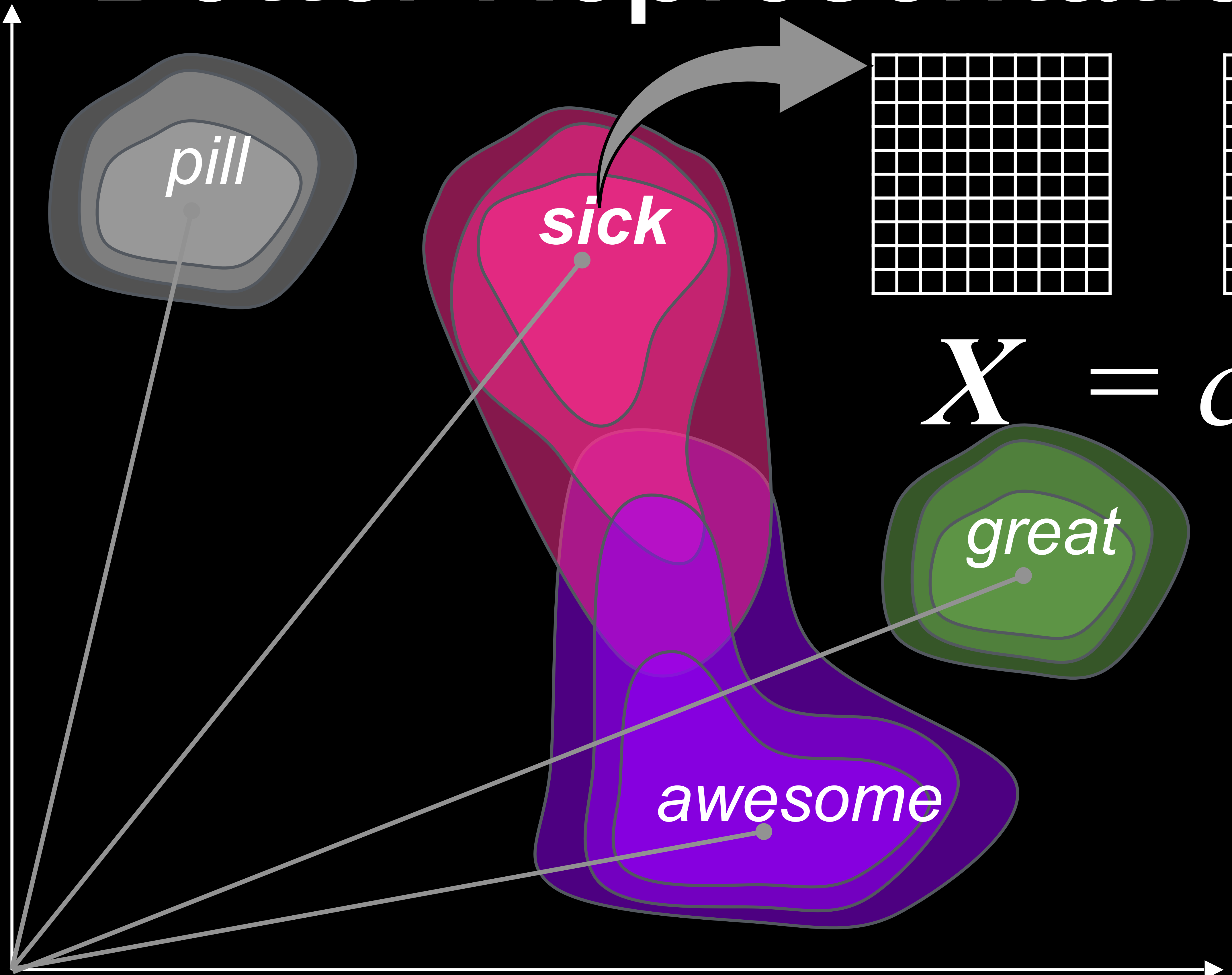


DIALECT →



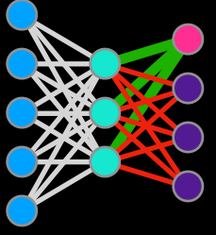
Bocconi

Better Representations



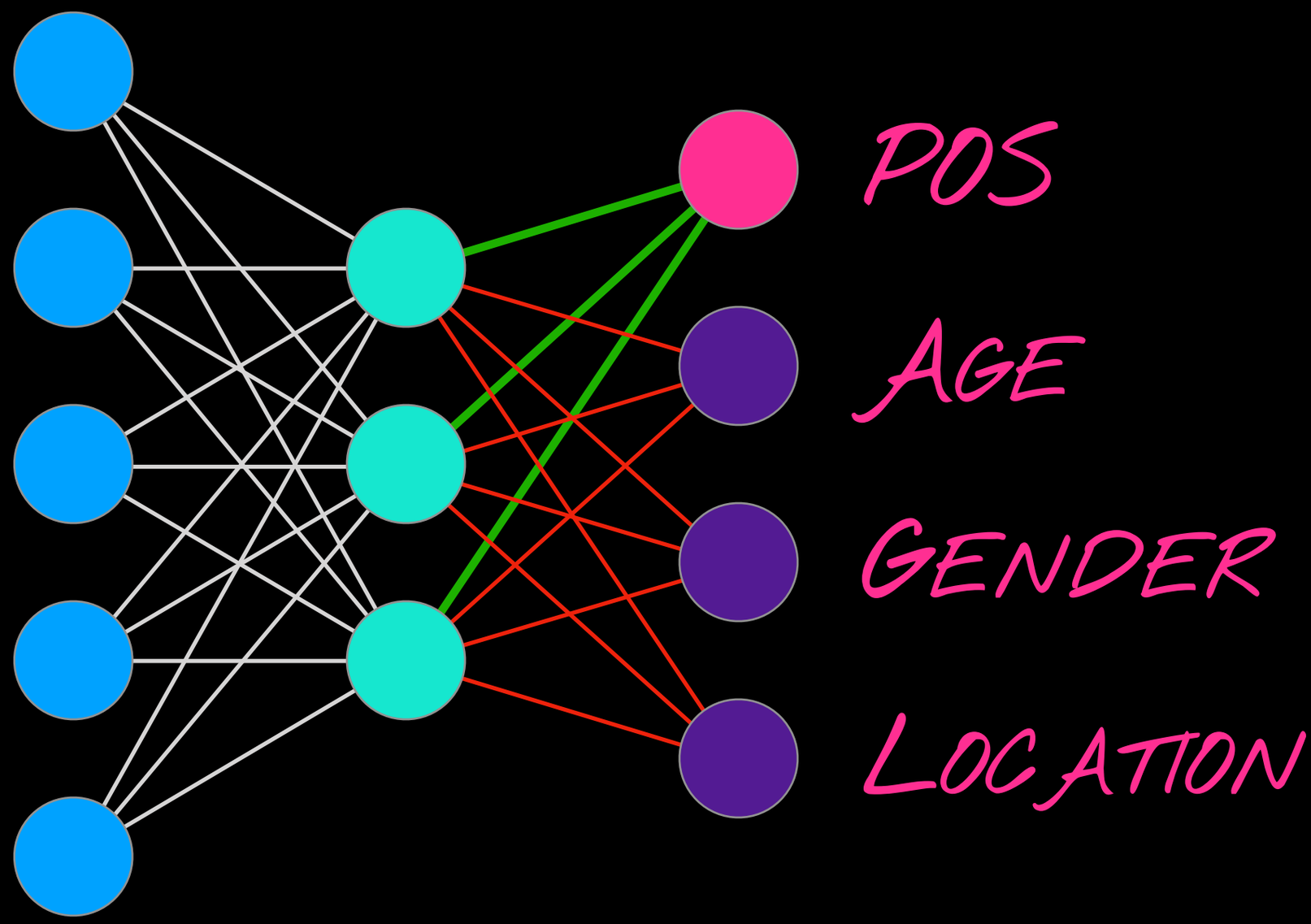
$$X = c_{ontext} \cdot s_{emantics}$$

 Bianchi, Attanasio, & Hovy, forthcoming



Better Models

- representation learning
- domain adaptation
- multitask learning
- reinforcement learning



45 Benton, Mitchell & Hovy, 2017; Li et al. 2018; Fornaciari et al., 2021, Attanasio et al. forthcoming

Outcomes

society:

combat algorithmic racism and sexism,

build fair tools that perform equally well for all users

research:

open up new research avenues and subfields

industry:

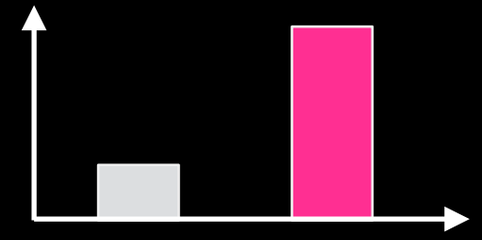
more performant tools in MT, dialogue, search

Wrapping Up

Bocconi

Social Factor Goals

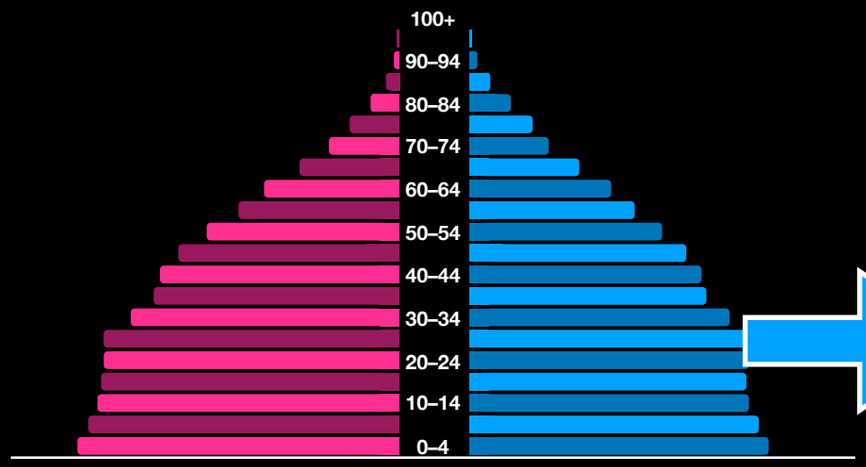
- develop novel methods to include social factors into NLP



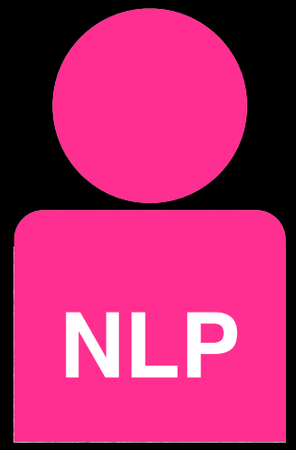
- improve existing performance

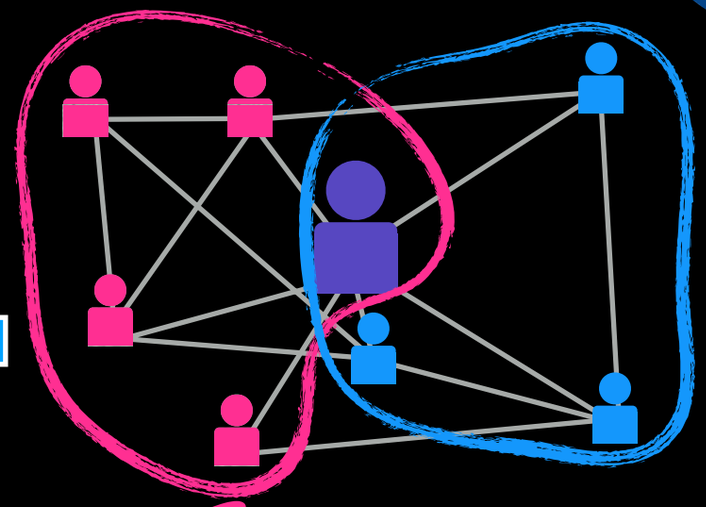
- enable novel personalized applications

Take-Home Points



DEMOGRAPHIC





SOCIAL

- People use language for information ***and*** social purposes
- NLP has modeled only information
- We are missing out on performance, personalization, fairness
- Let's model social aspects of language!

Bocconi

www.dirkhovy.com/portfolio/papers
milanlproc.github.io/

Thank you!

IF YOU ARE INTERESTED IN WORKING WITH ME: I HAVE UPCOMING POSTDOC POSITIONS IN THESE AREAS

@dirk_hovy
www.dirkhovy.com

INTEGRATOR
European Research Council

Bocconi

www.dirkhovy.com/portfolio/papers
milanlproc.github.io/

Questions?

IF YOU ARE INTERESTED IN WORKING WITH ME: I HAVE UPCOMING POSTDOC POSITIONS IN THESE AREAS

@dirk_hovy
www.dirkhovy.com

INTEGRATOR
European Research Council

Bocconi