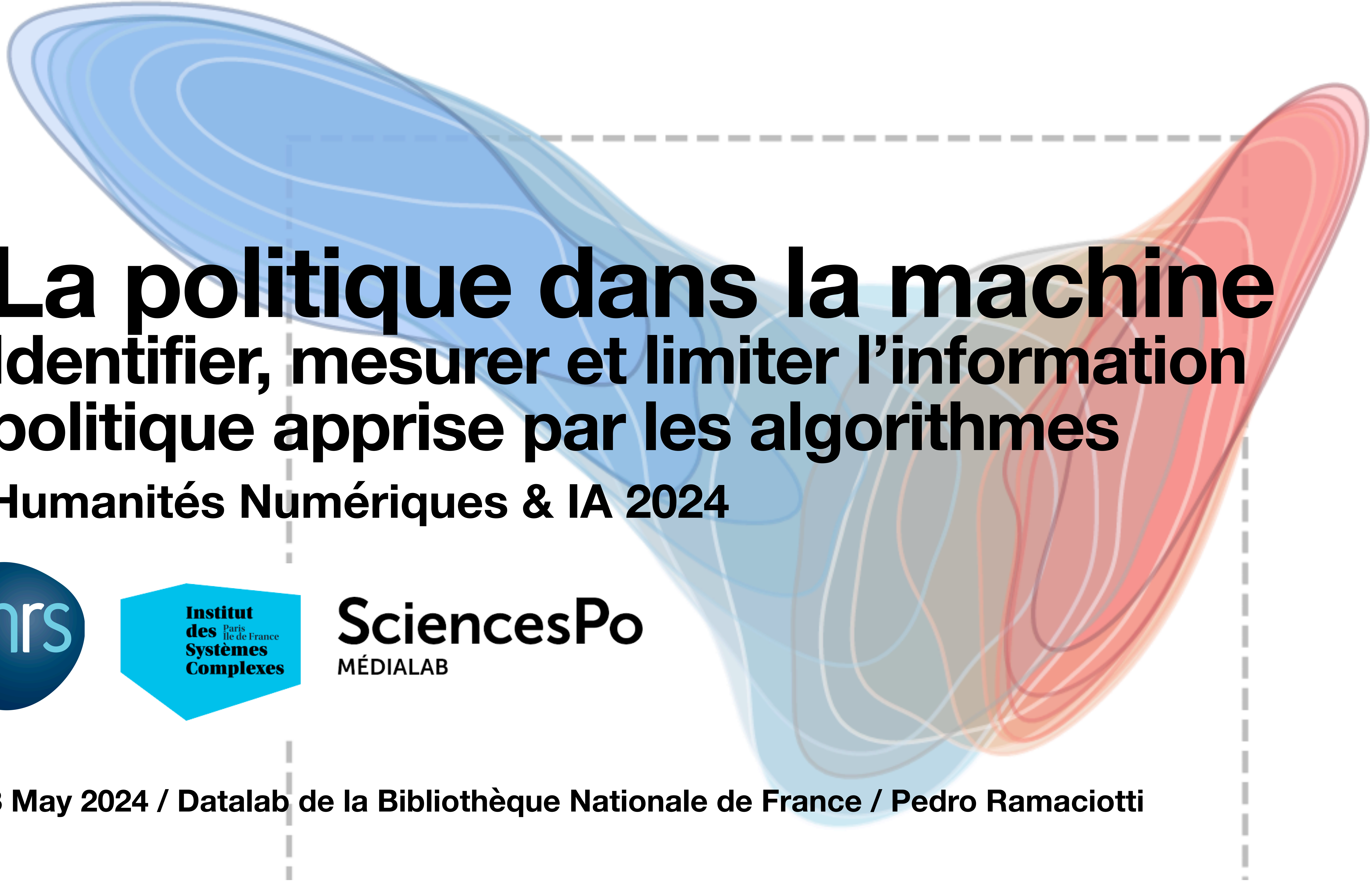


La politique dans la machine

Identifier, mesurer et limiter l’information politique apprise par les algorithmes

Humanités Numériques & IA 2024

3 May 2024 / Datalab de la Bibliothèque Nationale de France / Pedro Ramaciotti

Research program (in a nutshell)

1. Use individual **political position estimates** in large online populations to study **recommender systems**

2. **Audit** algorithmic outcomes (i.e, what's recommended), but also advance **AI explainability hinging on politics**

3. Develop **toolkits for algorithm design, compliance and platform regulation**

EPO

European Polarisation Observatory

cnrs

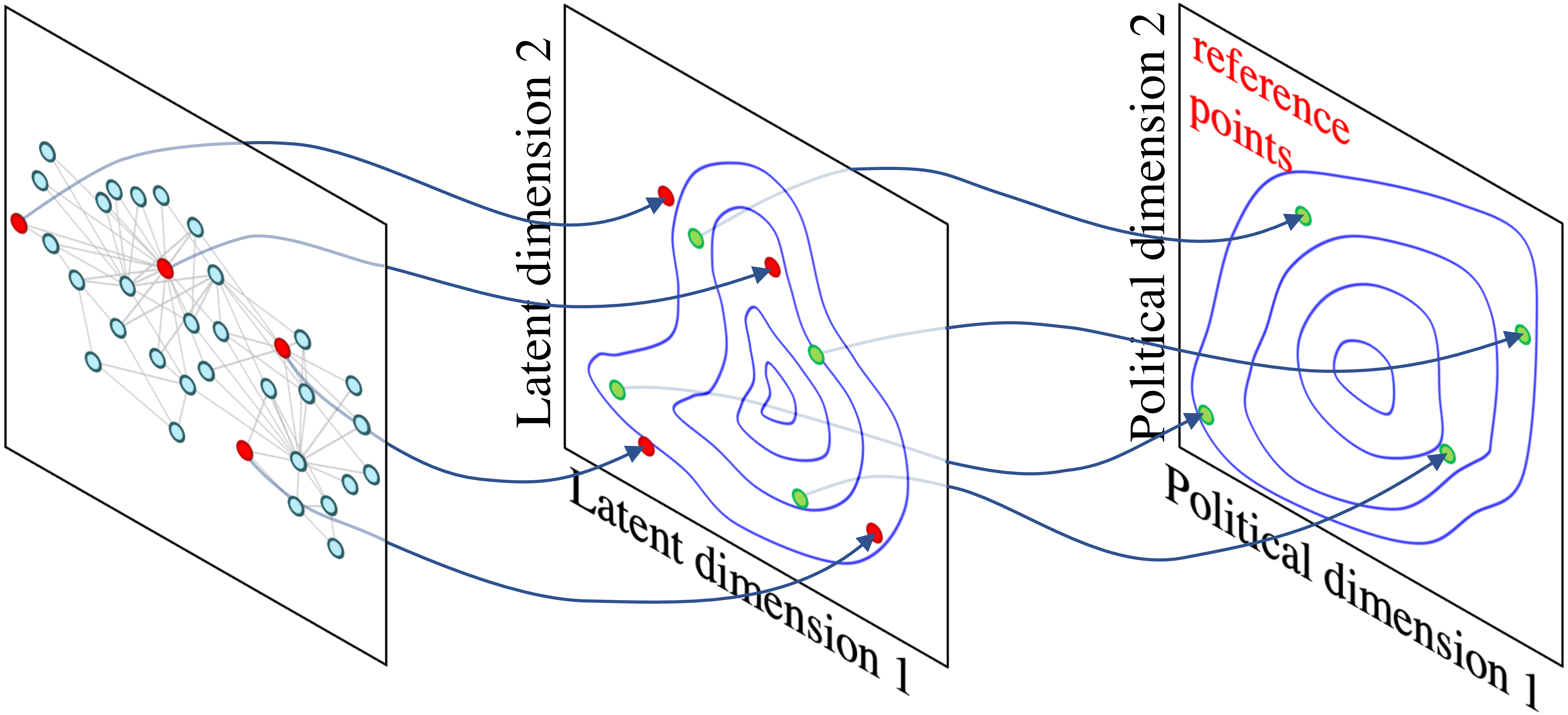
SciencesPo

CEU CENTRAL EUROPEAN UNIVERSITY

Università Bocconi MILANO

Multidimensional Political Opinions Spaces

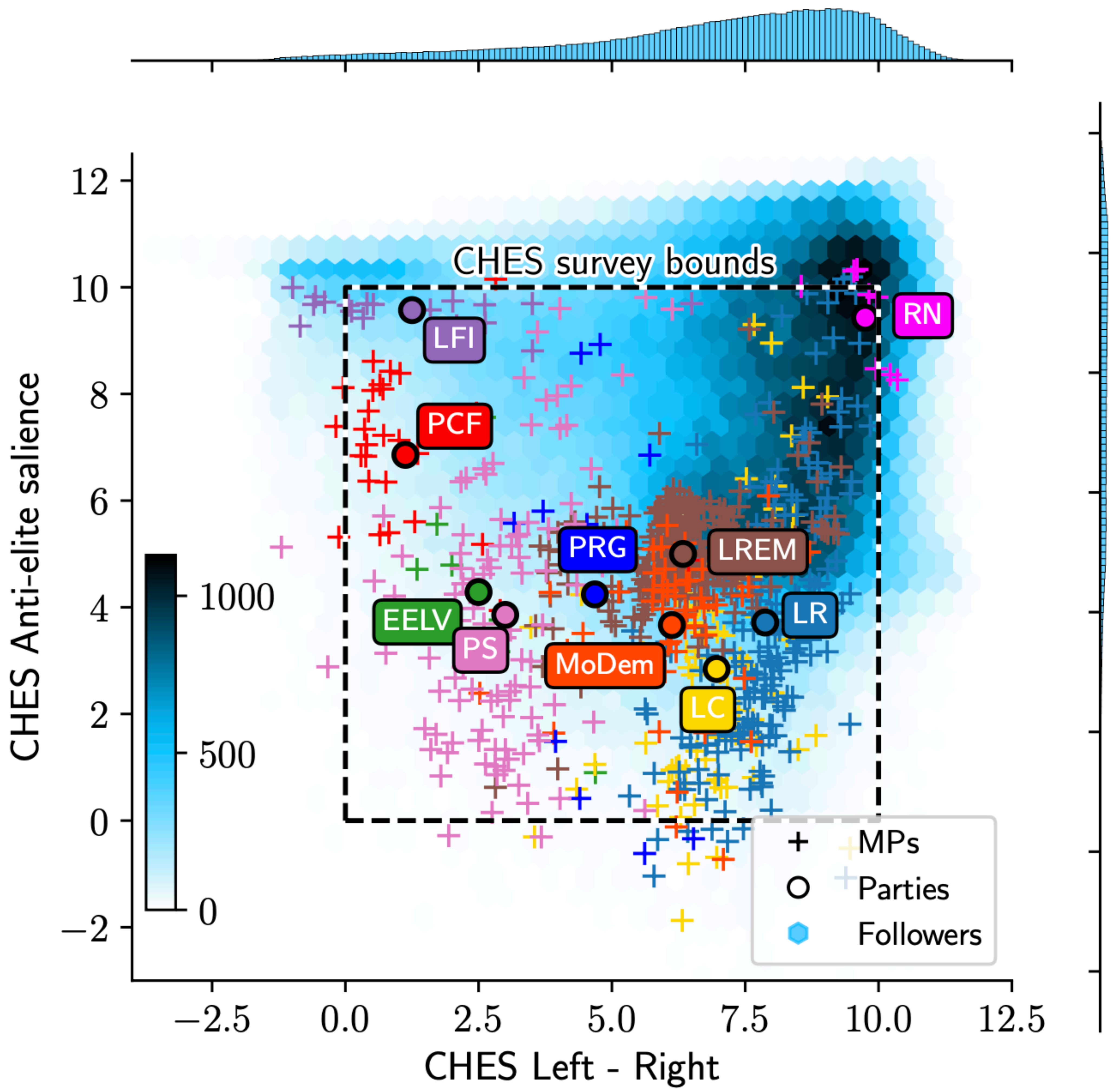
Embedding procedure

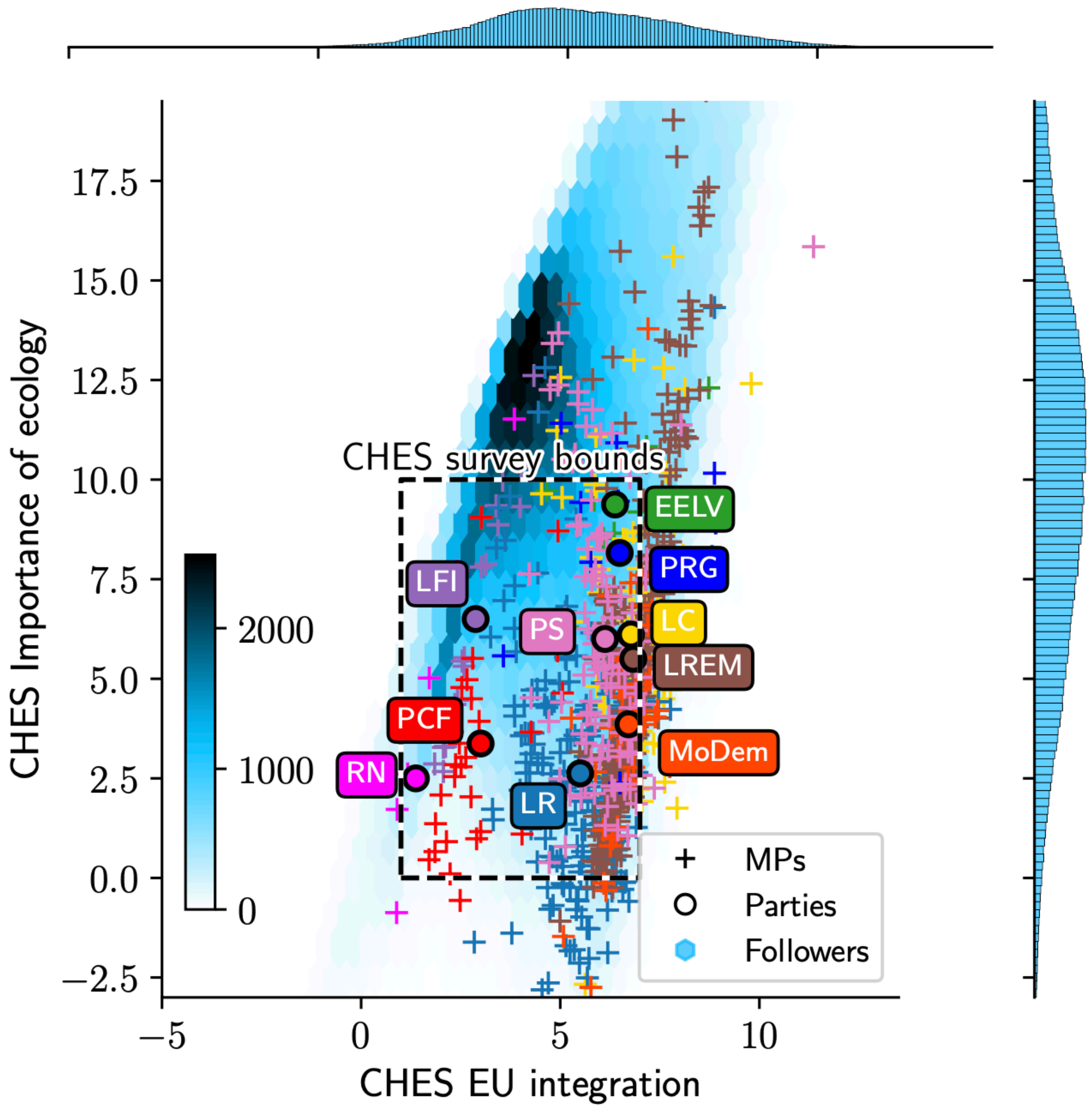


Social graph ⇨ Latent space ⇨ CHES space

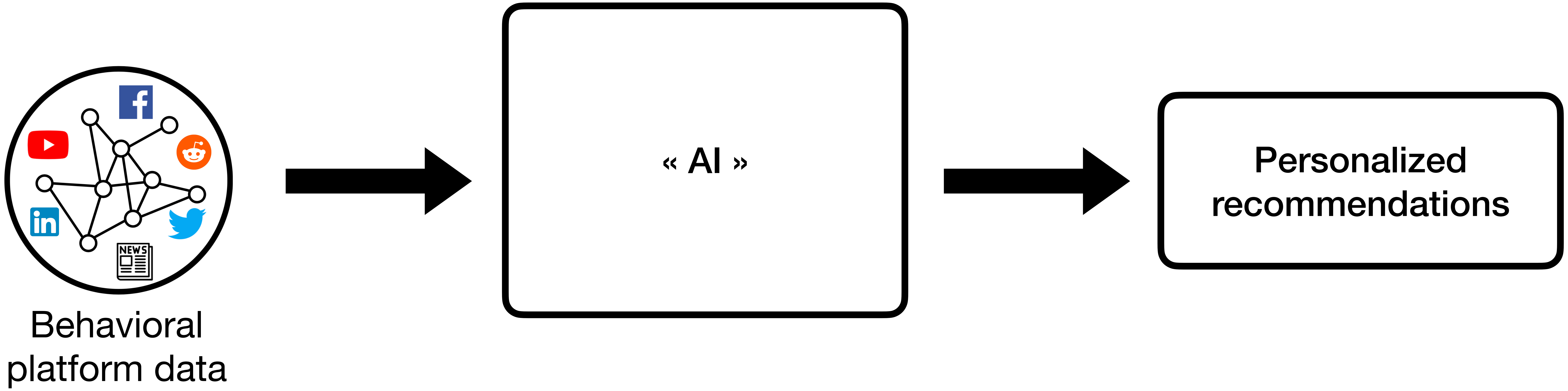
Multidimensional Political Opinions Spaces

Embedding procedure





AI explainability and online politics

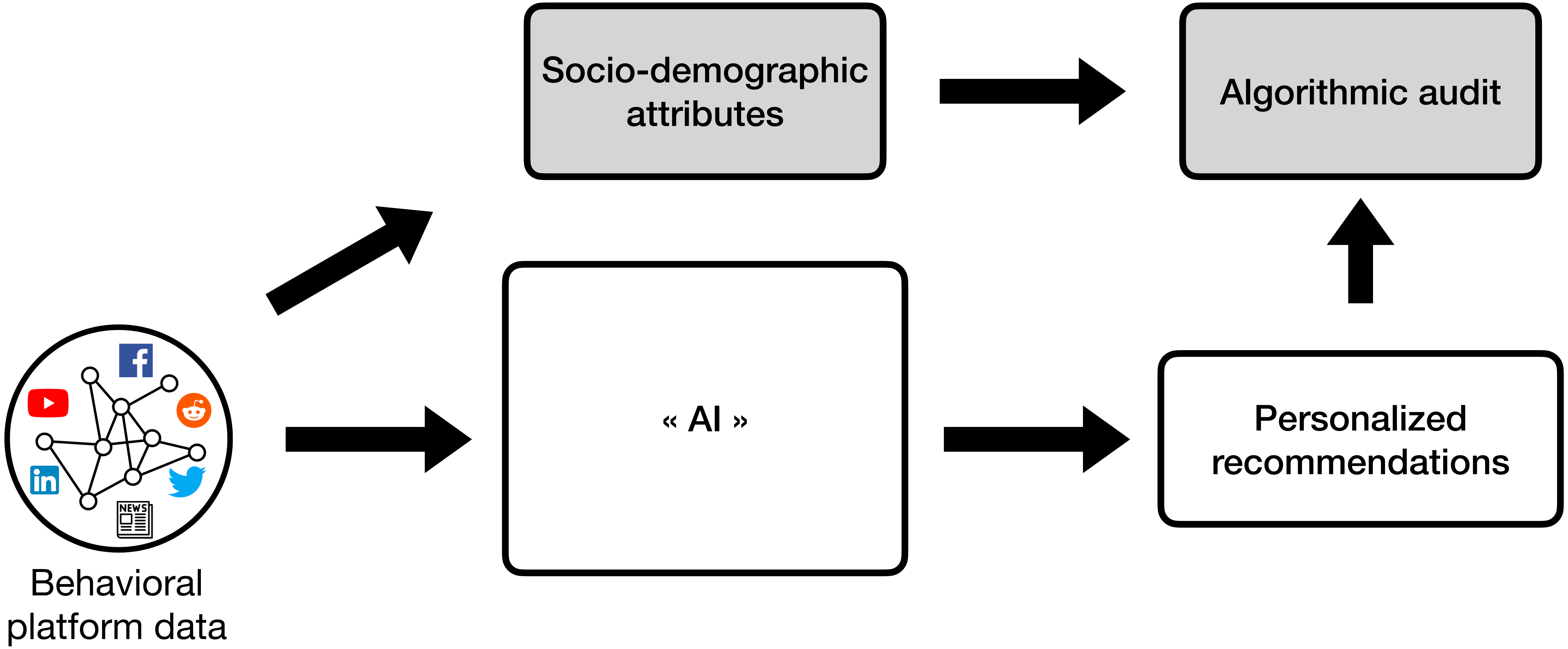


Behavioral platform data

« AI »

Personalized recommendations

AI explainability and online politics



Socio-demographic attributes

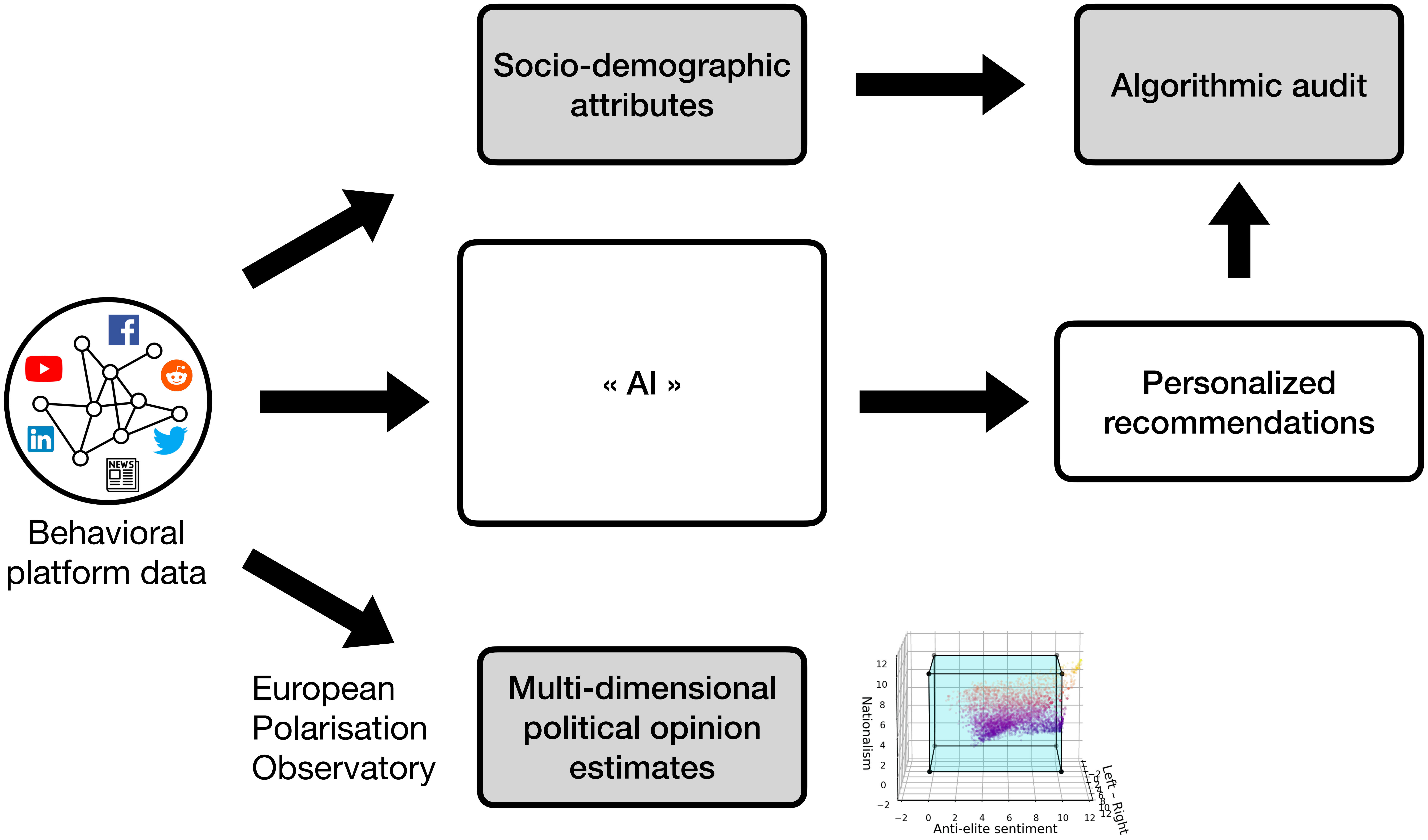
Algorithmic audit

« AI »

Personalized recommendations

Behavioral platform data

AI explainability and online politics



Behavioral platform data

Socio-demographic attributes

« AI »

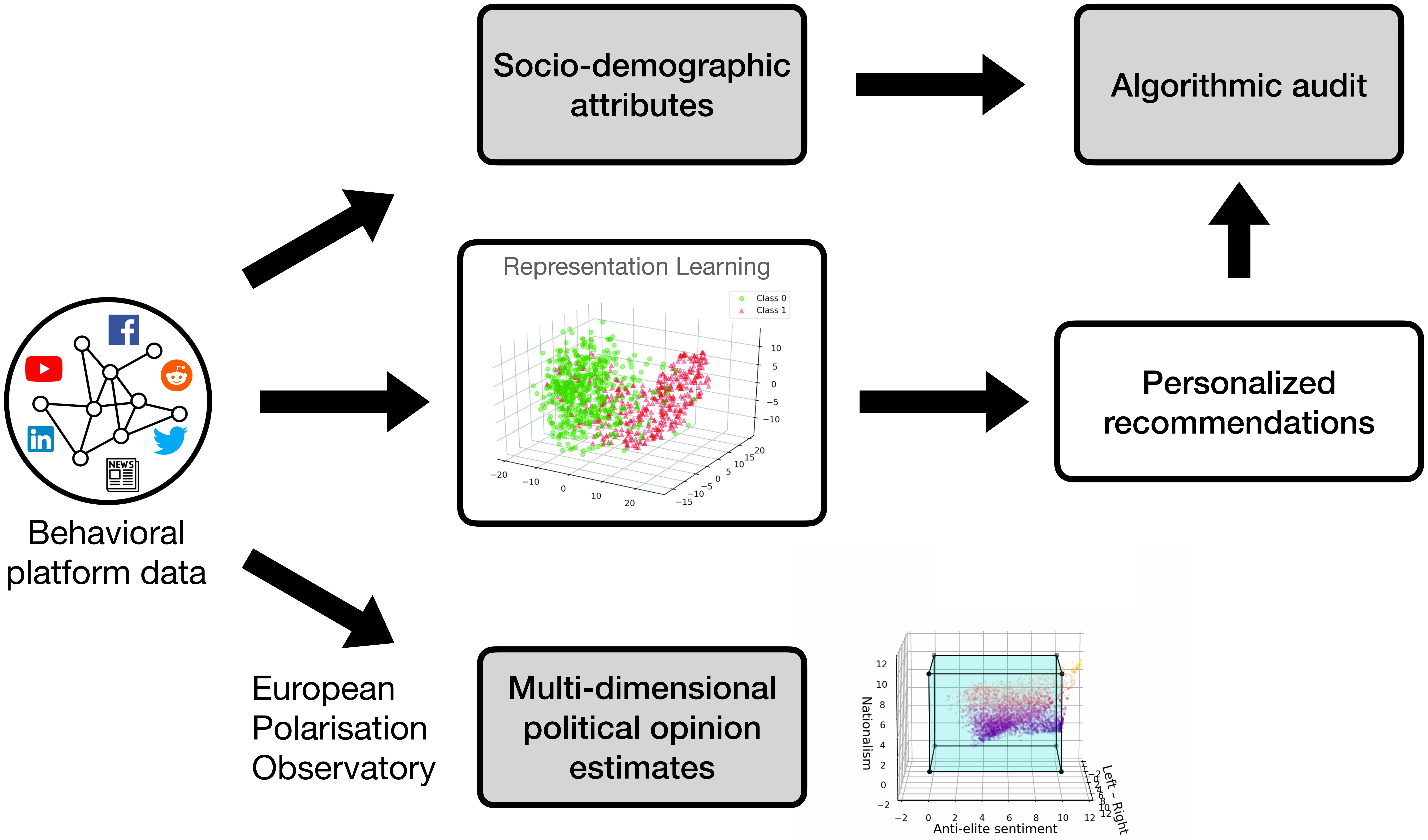
Multi-dimensional political opinion estimates

European Polarisation Observatory

Algorithmic audit

Personalized recommendations

AI explainability and online politics



Behavioral platform data

Socio-demographic attributes

Algorithmic audit

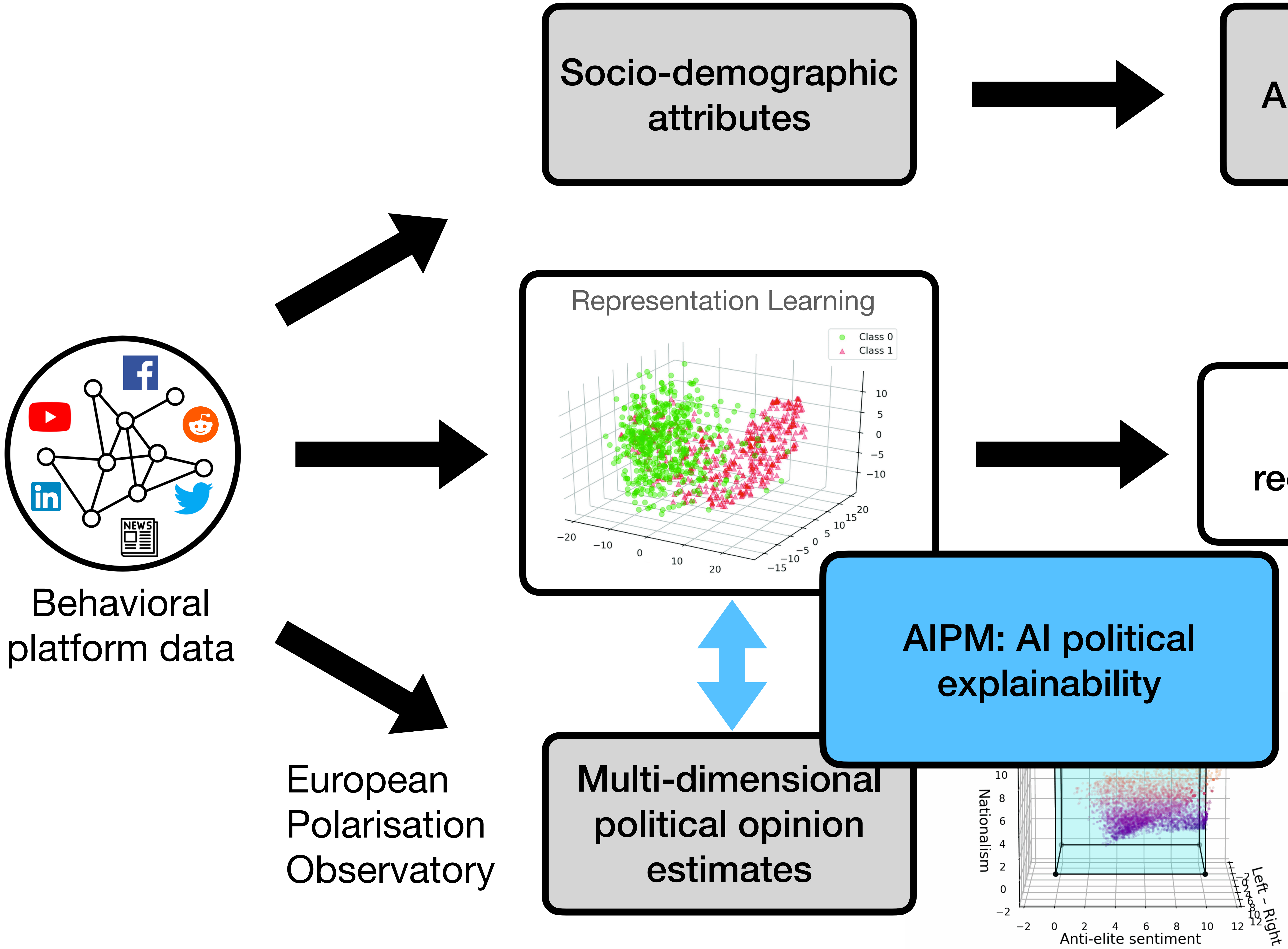
Representation Learning

Personalized recommendations

European Polarisation Observatory

Multi-dimensional political opinion estimates

AI explainability and online politics



Behavioral platform data

Socio-demographic attributes

Algorithmic audit

Representation Learning

Personalized recommendations

AIPM: AI political explainability

European Polarisation Observatory

Multi-dimensional political opinion estimates

First study: audit in opinion spaces

- Collaboration with Horus project at CNRS
  - Panel of 2.258 users that volunteer platform data (browser plug-in)
  - 17M recommendations (+engagement) on Twitter (Oct 2022 – Jan 2024)

First study: audit in opinion spaces

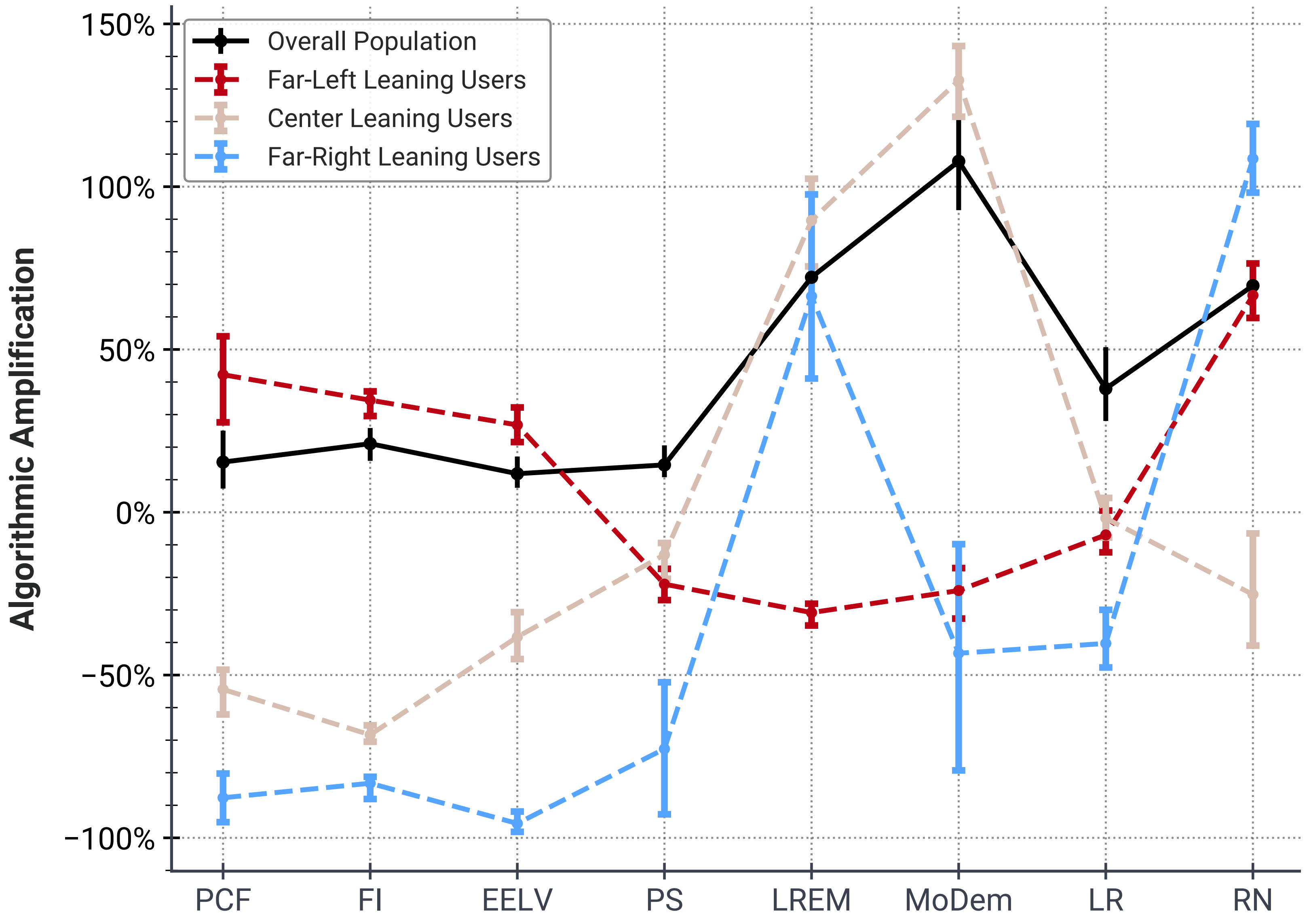
- Collaboration with Horus project at CNRS
  - Panel of 2.258 users that volunteer platform data (browser plug-in)
  - 17M recommendations (+engagement) on Twitter (Oct 2022 – Jan 2024)
- Collected tweets posted by friends of panel
  - Caveat: we cannot address content-based recommendations of global scope

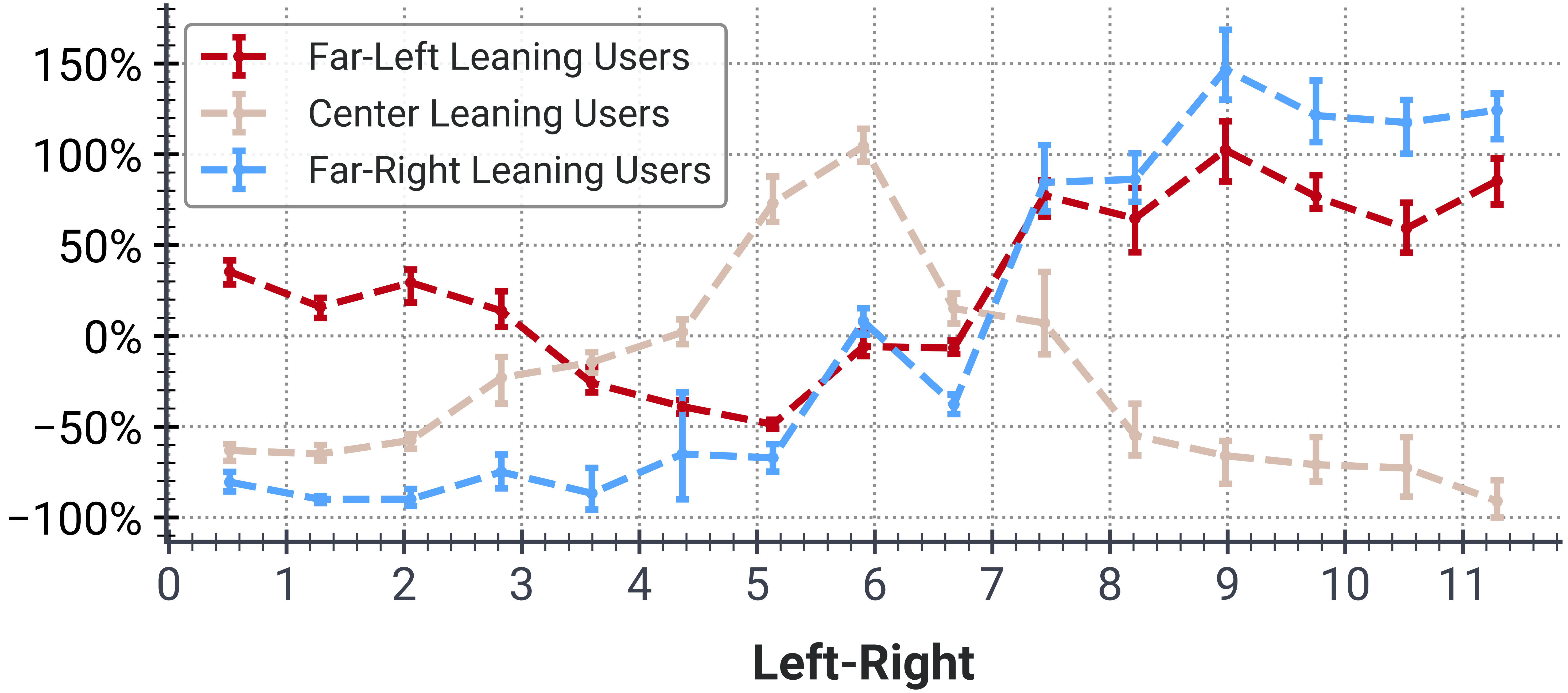
First study: audit in opinion spaces

- Collaboration with Horus project at CNRS
  - Panel of 2.258 users that volunteer platform data (browser plug-in)
  - 17M recommendations (+engagement) on Twitter (Oct 2022 – Jan 2024)
- Collected tweets posted by friends of panel
  - Caveat: we cannot address content-based recommendations of global scope
- Measured algorithmic amplification with respect to time-reversal recommender

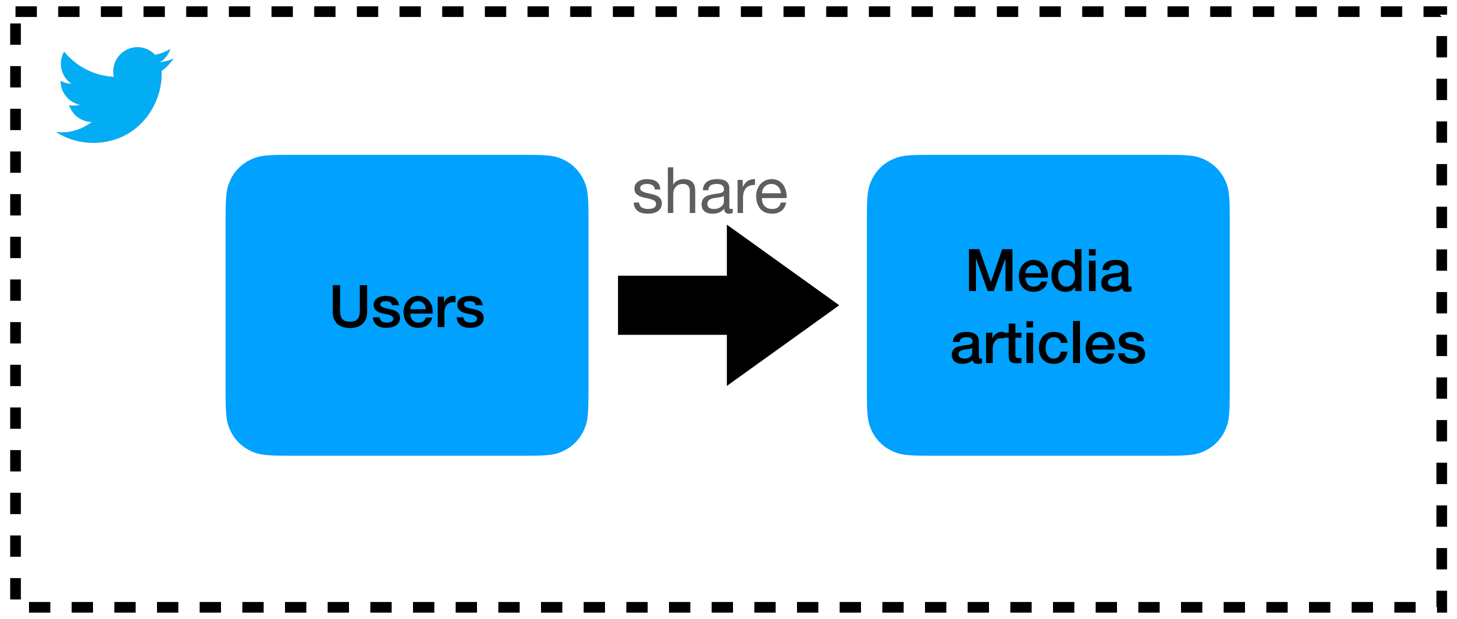
First study: audit in opinion spaces

Algorithmic asymmetries in French Twitter

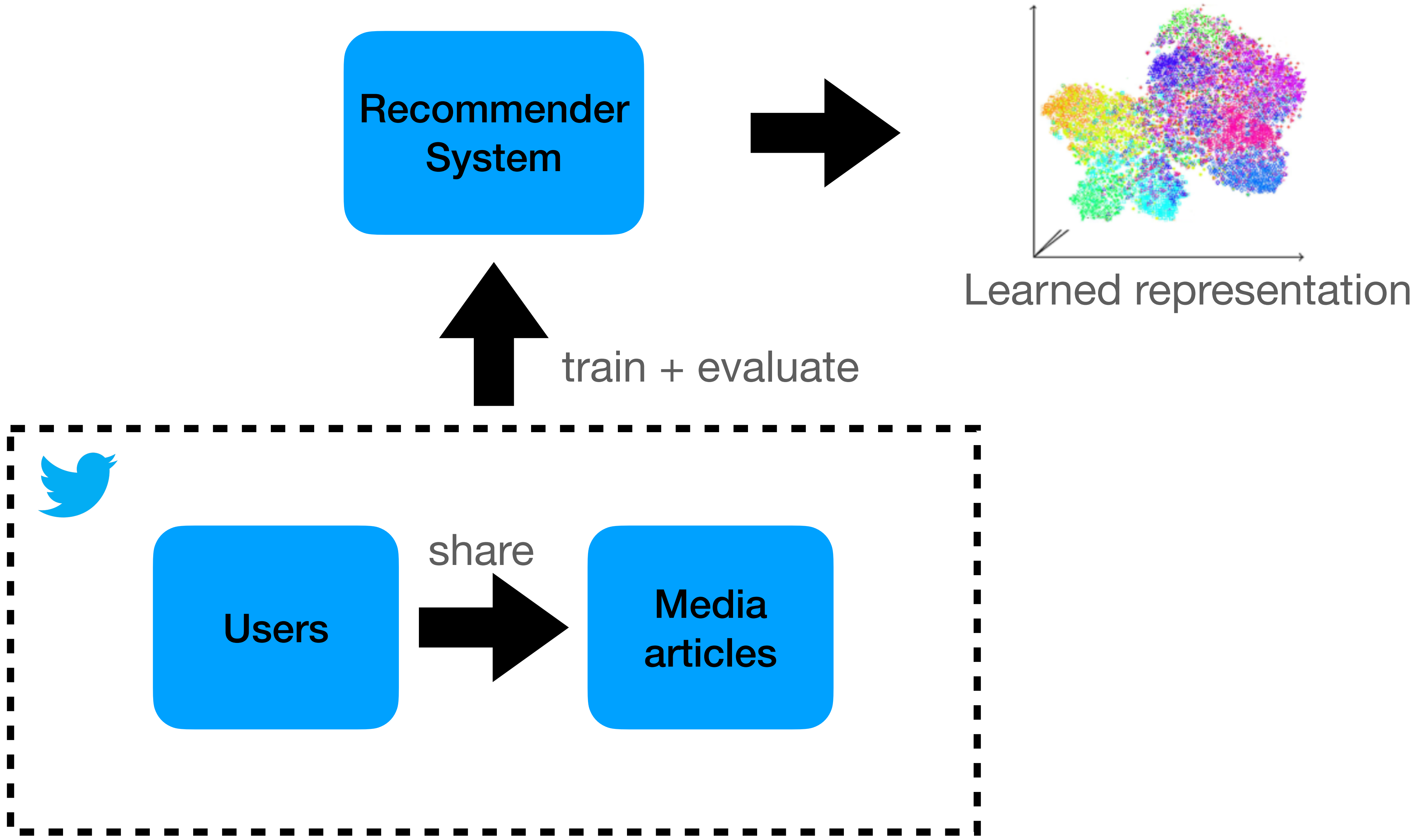




Second study: breaking AI political explainability



Second study: breaking AI political explainability



Recommender System

Learned representation

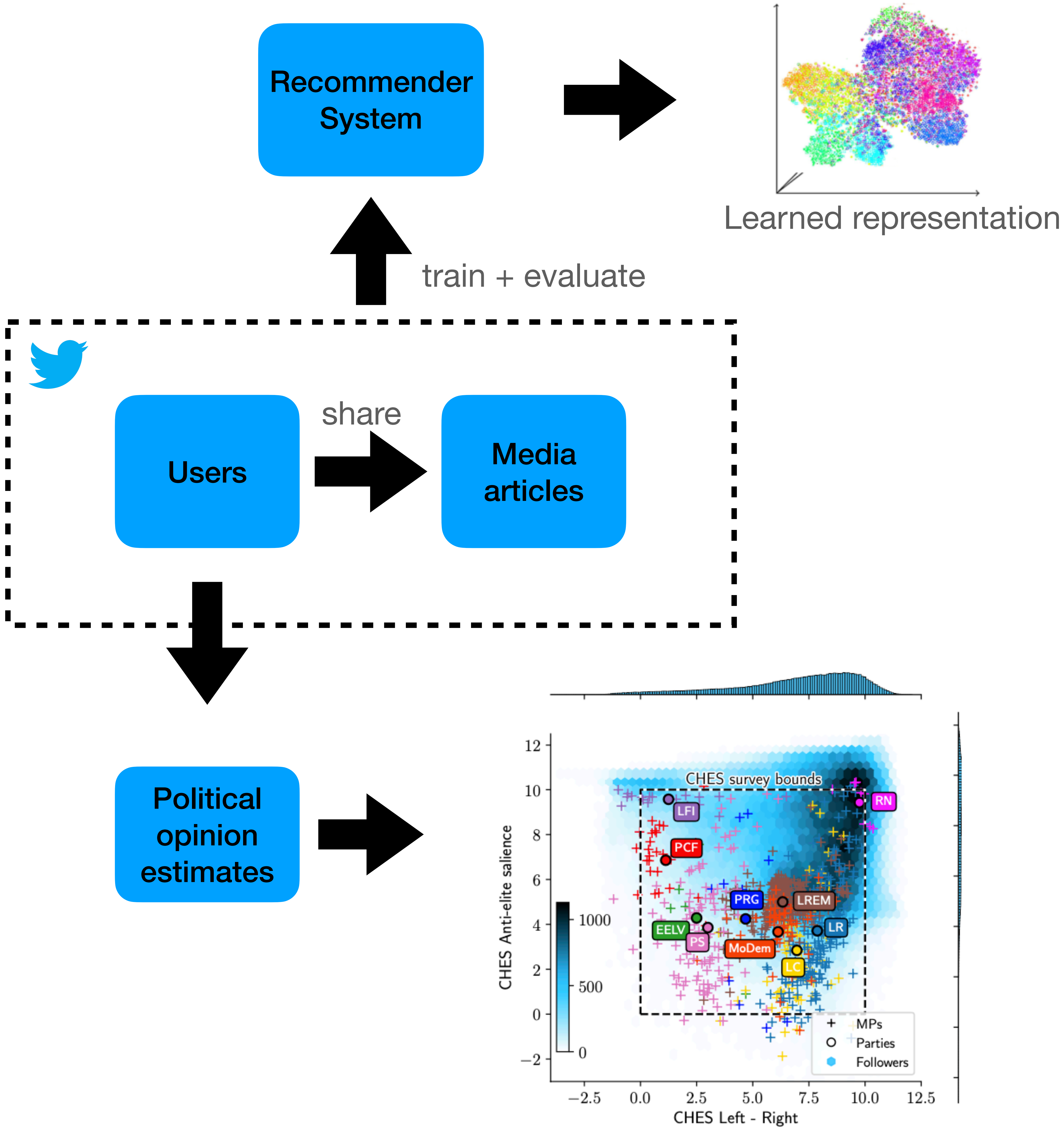
train + evaluate

Users

share

Media articles

Second study: breaking AI political explainability



Recommender System

Learned representation

train + evaluate

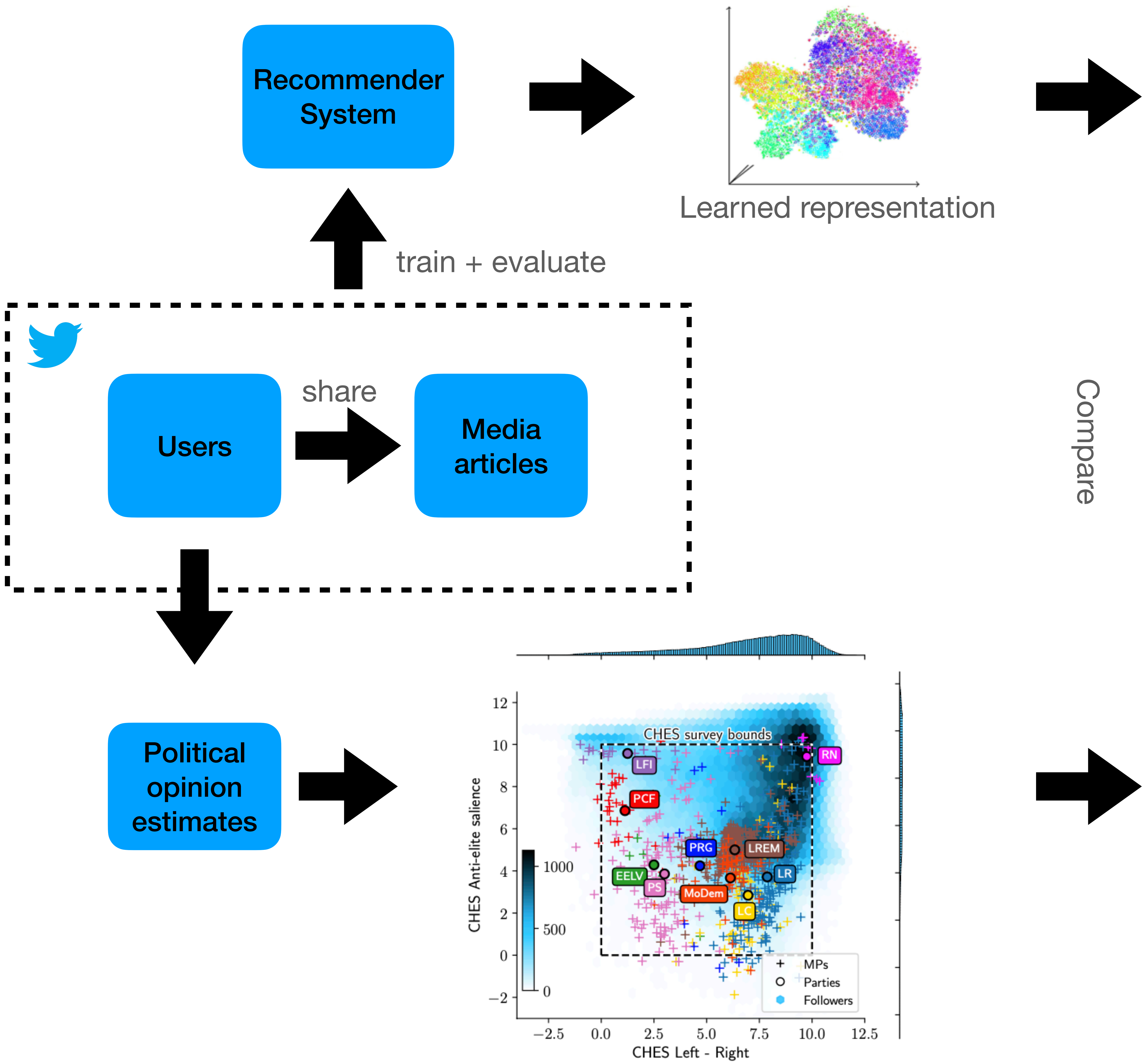
Users

share

Media articles

Political opinion estimates

Second study: breaking AI political explainability



Recommender System

Learned representation

train + evaluate

Users

share

Media articles

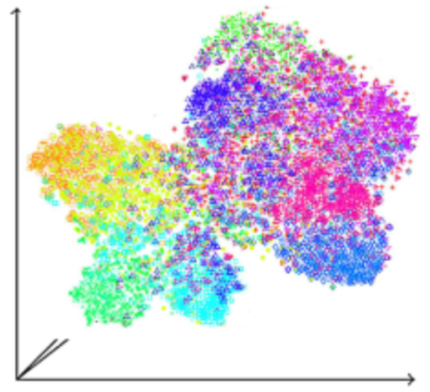
Political opinion estimates

Compare

Second study: breaking AI political explainability

Recommender System

train + evaluate



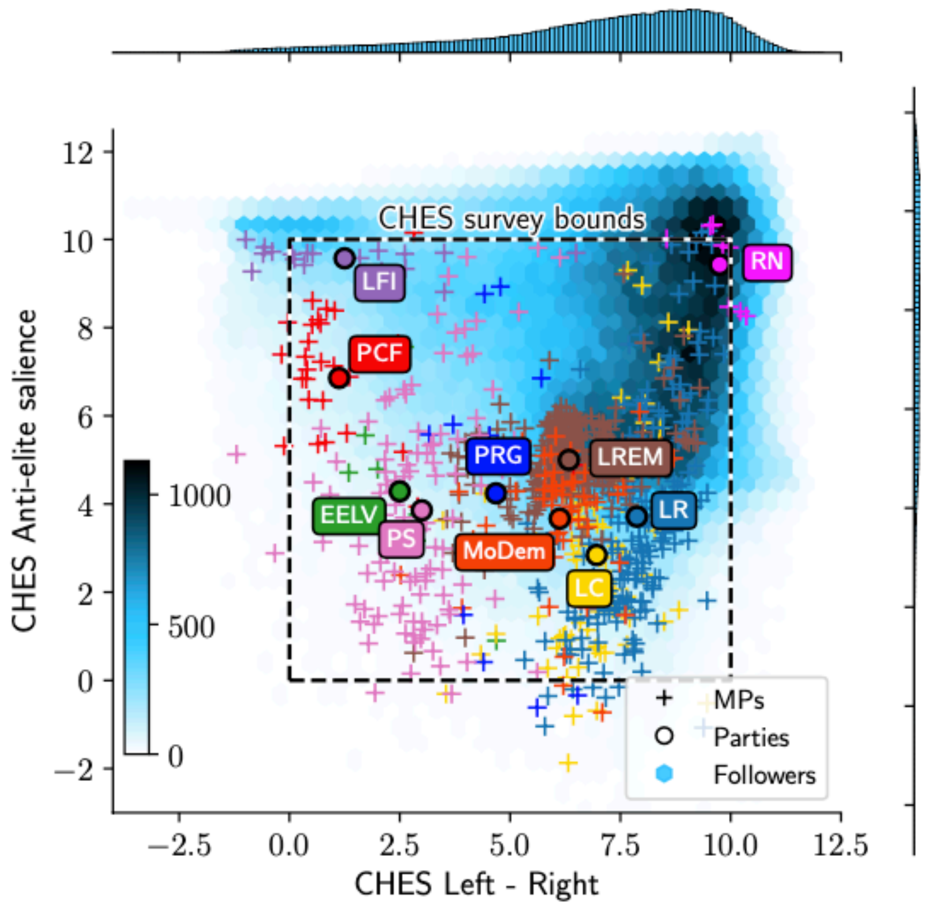
Learned representation

Users

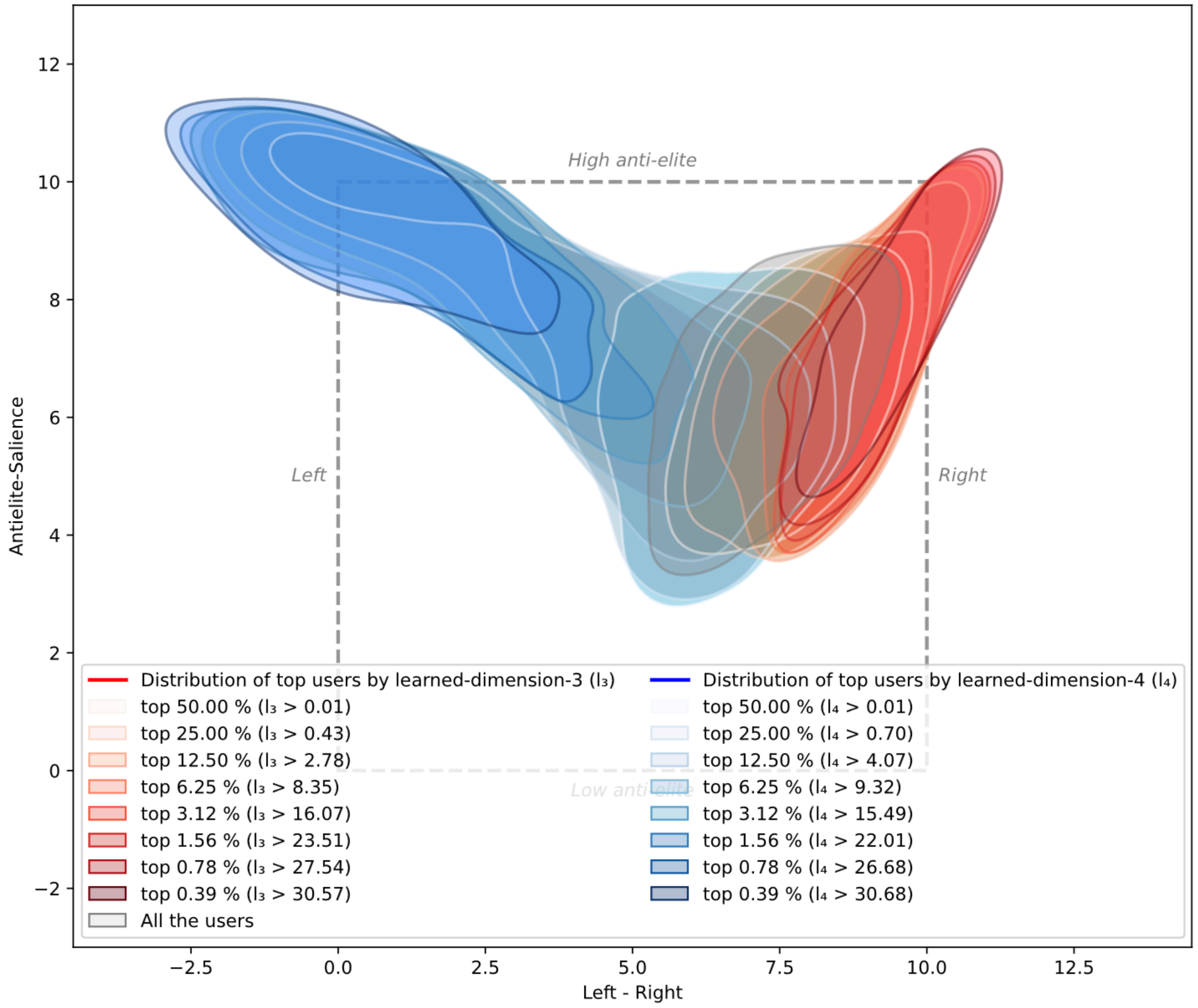
share

Media articles

Political opinion estimates



Compare



Toolkits, compliance & platform regulation

Regulation, compliance and toolkits

- GDPR, Art 9:

  Processing of data containing political opinions (among other « special » sensitive categories) shall be prohibited.

Regulation, compliance and toolkits

- GDPR, Art 9:

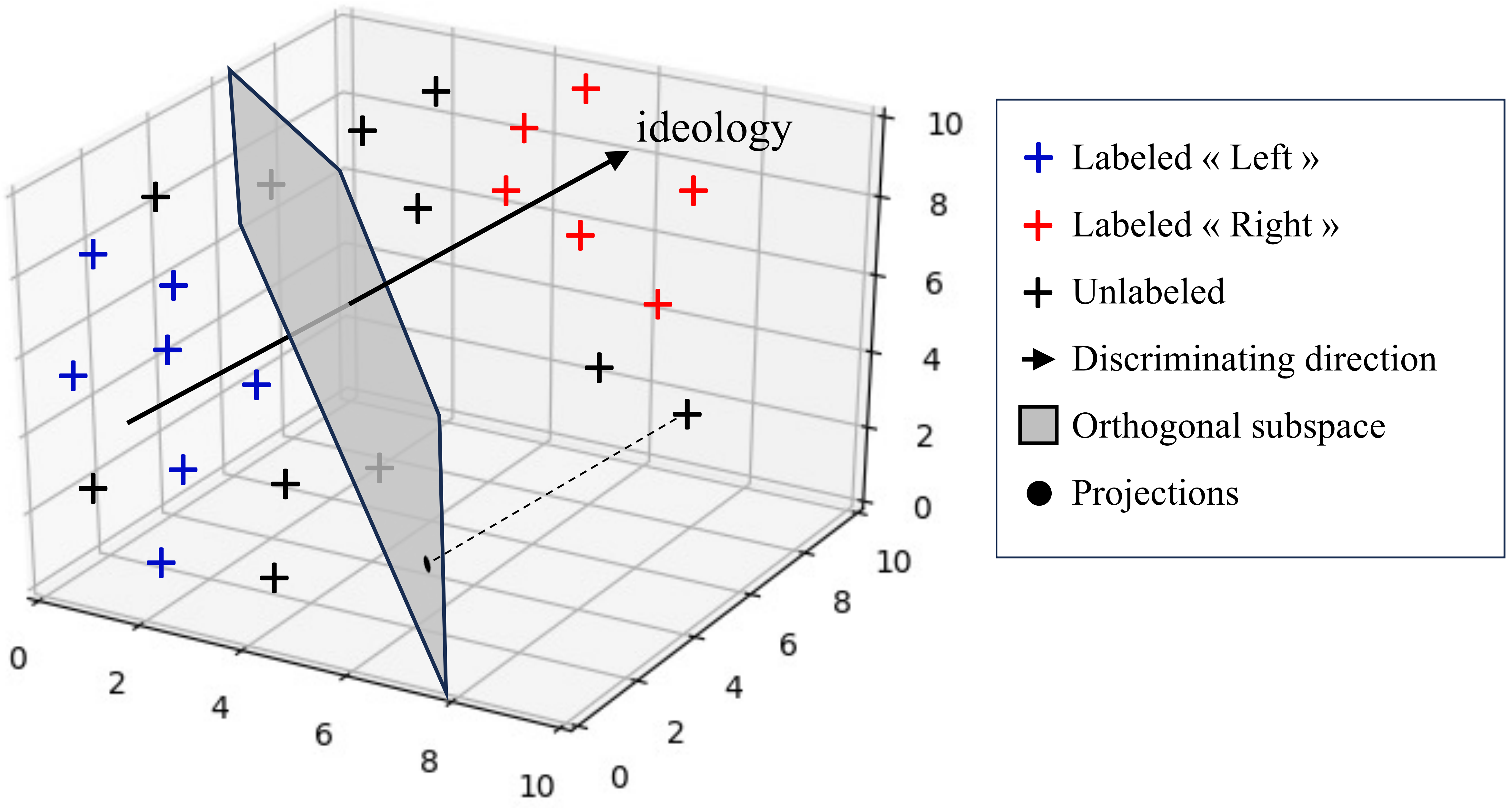
  Processing of data containing political opinions (among other « special » sensitive categories) shall be prohibited.

- DSA, Art 26:

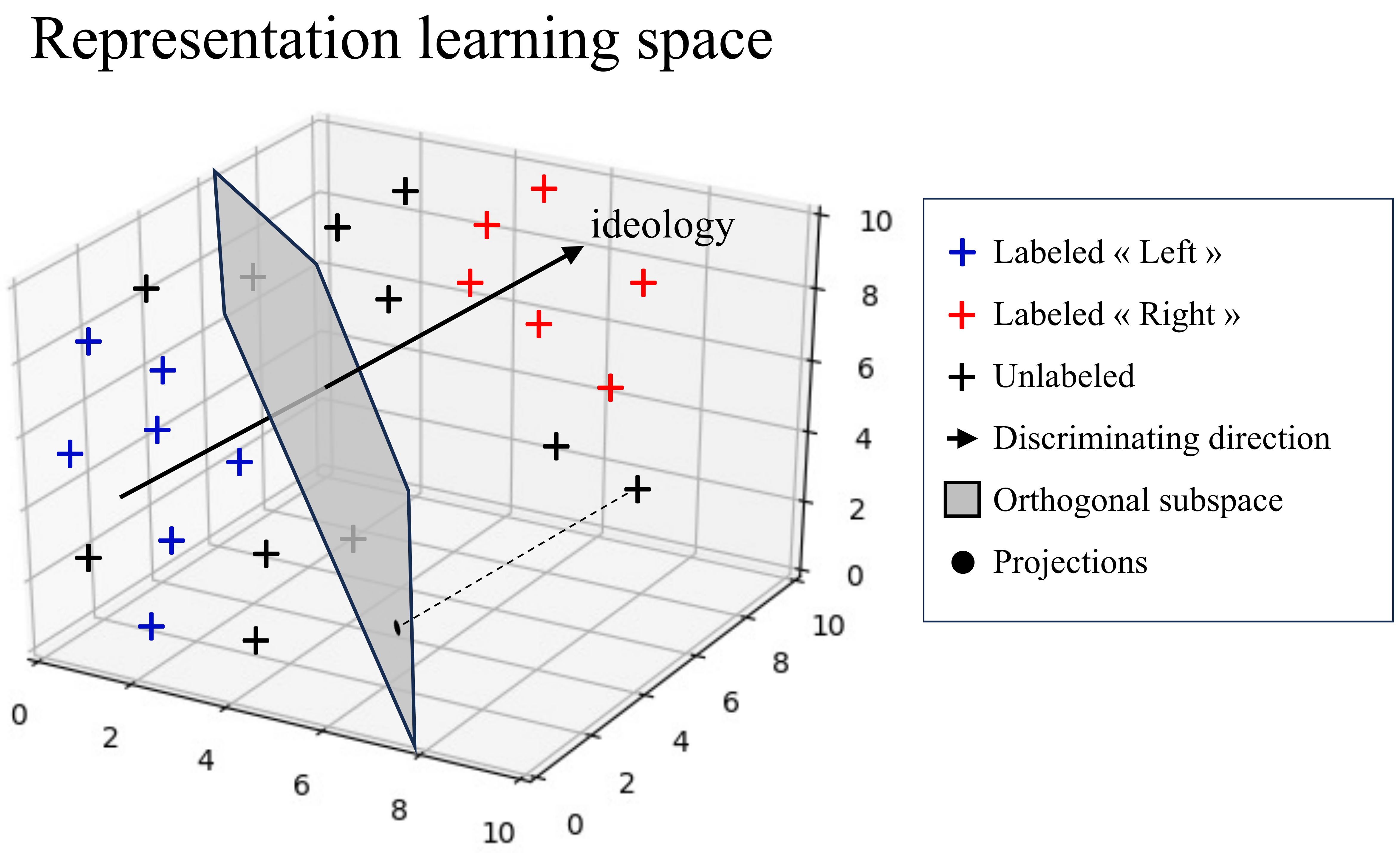
  Platforms shall not recommend ads based on profiling on sensitive categories under Art 9 of GDPR (including political opinions).

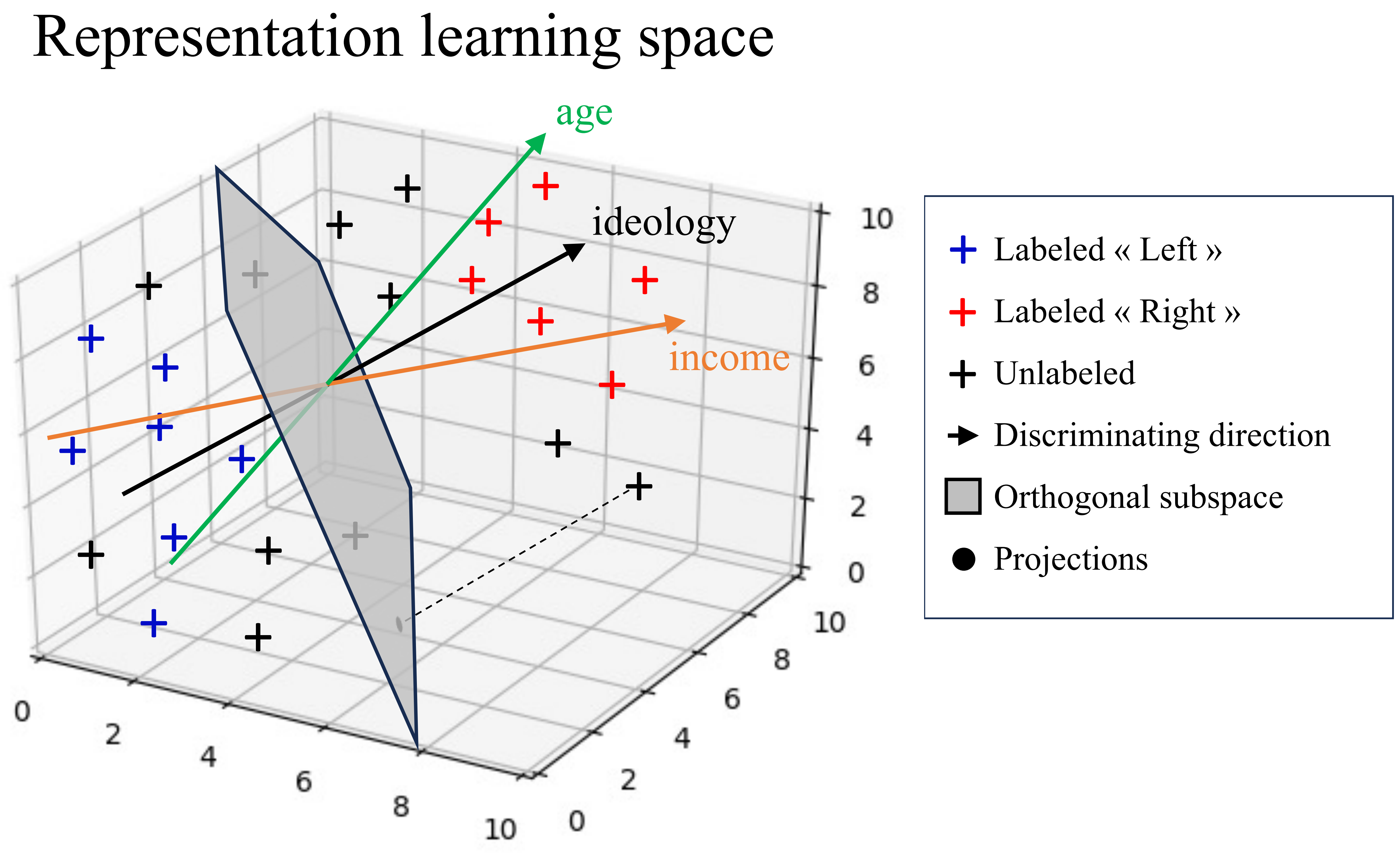
Regulation, compliance and toolkits

Representation learning space

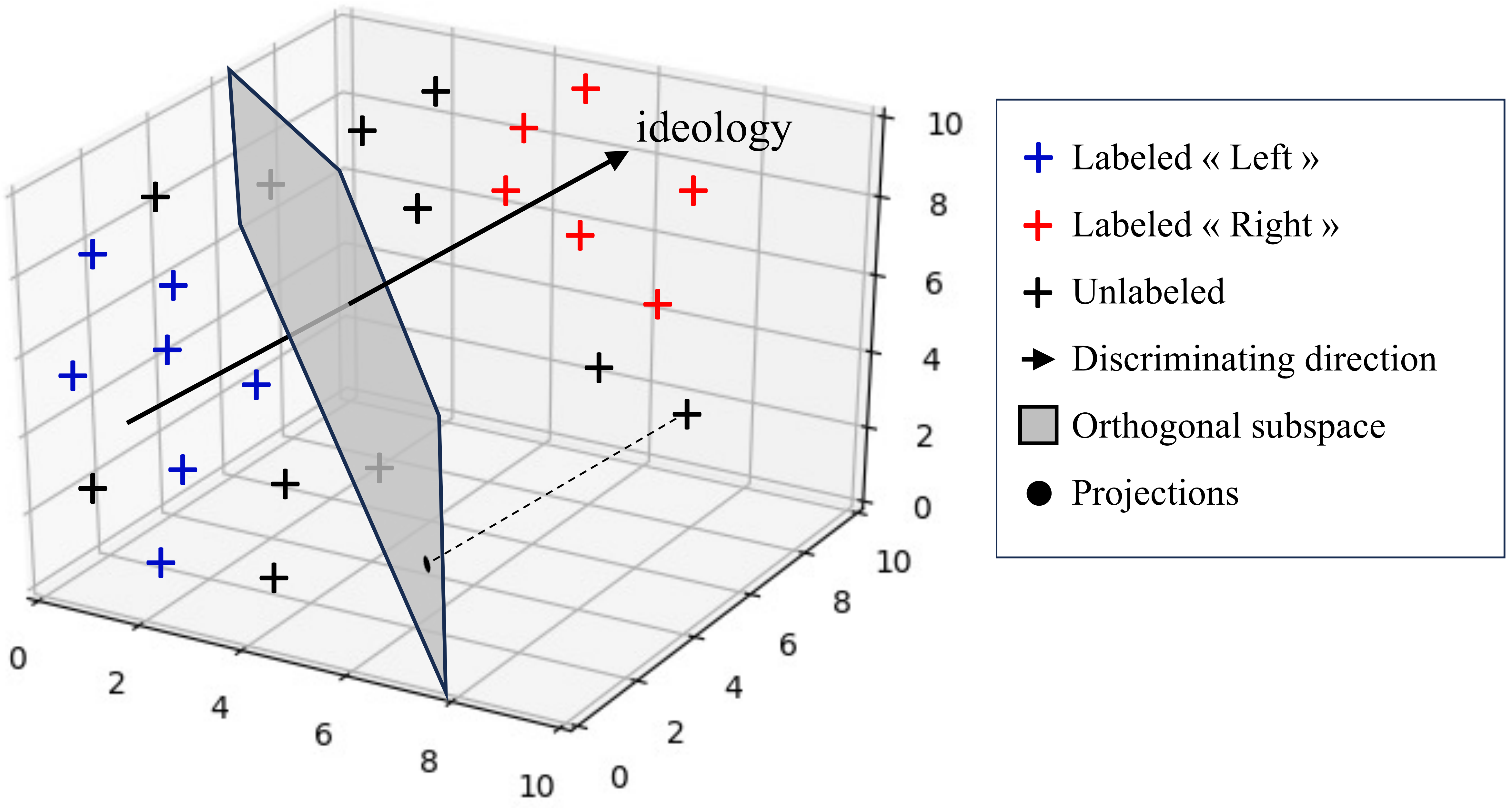


Regulation, compliance and toolkits

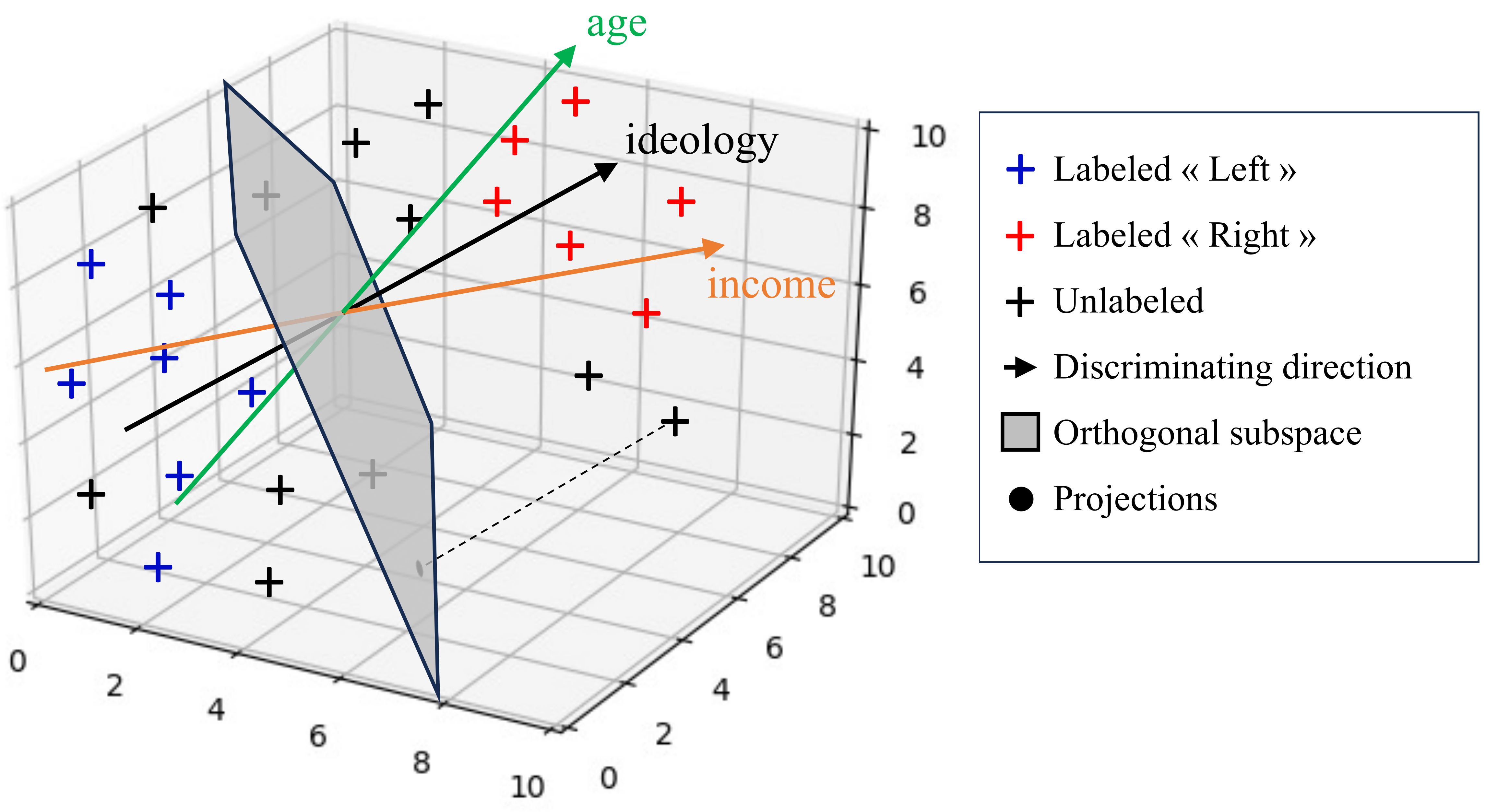




Regulation, compliance and toolkits



Representation learning space



Representation learning space

Profiling, political opinion, compliance with GDPR/DSA…

Q&A