Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

HUMAN-AI COEVOLUTION

Towards a society centric approach

Dino Pedreschi, Luca Pappalardo, Emanuele Ferragina, Ricardo Baeza-Yates, Albert-Laszlo Barabasi, Frank Dignum, Virginia Dignum, Tina Eliassi-Rad, Fosca Giannotti, Janos Kertesz, Alistair Knott, Yannis Ioannidis, Paul Lukowicz, Andrea Passarella, Alex Sandy Pentland, John Shawe-Taylor, Alessandro Vespignani

Presenter: Valentina Pansanella, ISTI-CNR

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

HUMAN-AI COEVOLUTION

A perpetual, iterative process wherein both humans and learning algorithms evolve in tandem, each influencing the evolution of the other over time.

RECOMMENDERS

AI-based algorithms that **suggest** items or content based on users' preferences or specific requests

- They mediate, *through online platforms*, most of our actions by exerting instant influence over many specific choices
- Studying the role of recommenders within human-AI ecosystems constitutes a vantage point to analyse coevolution

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

THE FEEDBACK LOOP

Interactions between *users* and *recommenders* always generate a **feedback loop**

- Users' choices determine data on which recommenders are trained;
- The trained recommenders exert influence on users' choices
- Which affect the next round of training
- and so on….

UNINTENDED CONSEQUENCES

- Personalised recommendations on social media may artificially amplify **echo chambers**, **filter bubbles**, and **radicalisation**
- Profiling and targeted advertising may increase **inequality** and monopolies, accruing **biases** and **discriminations**
- Navigation services suggest routes that may create **congestion** if too many drivers are sent to the same roads

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

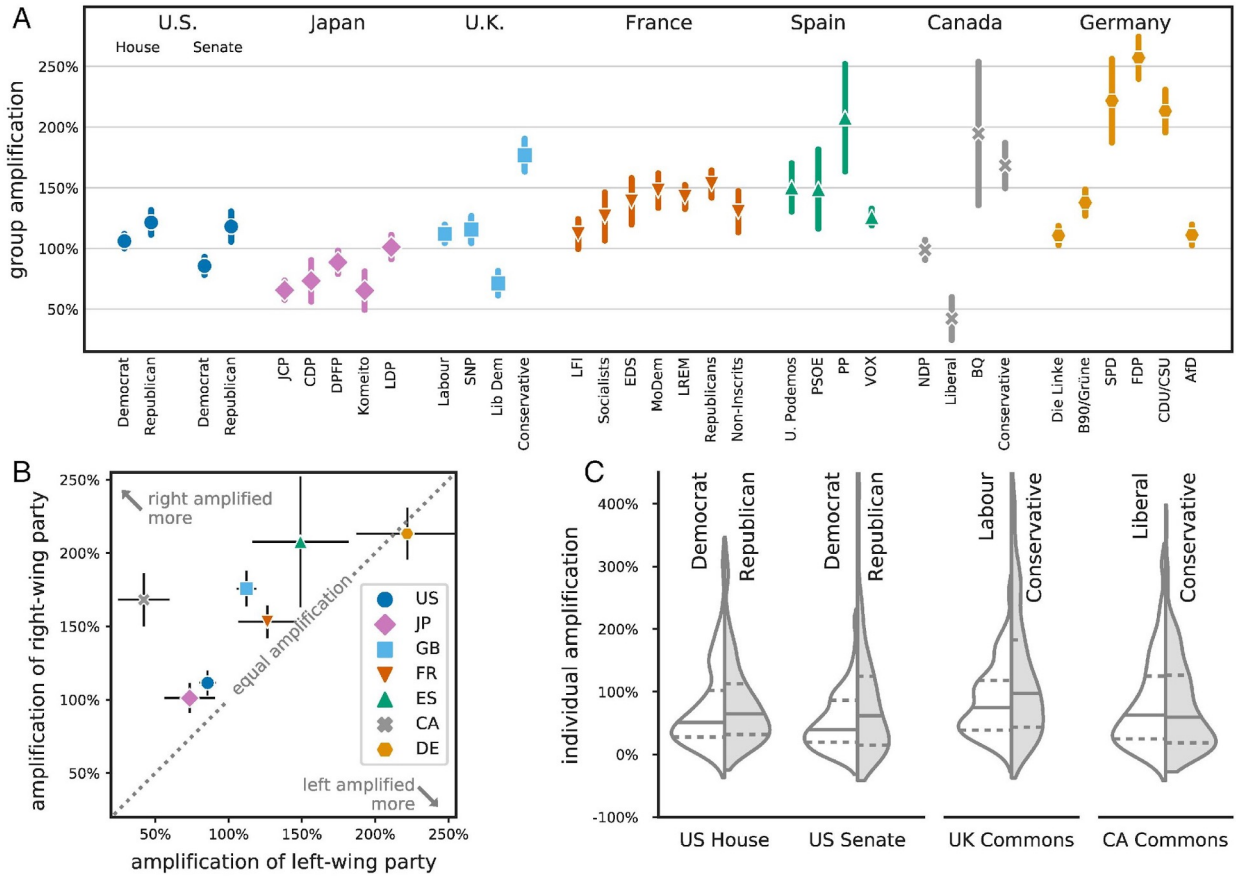
ALGORITHMIC BIAS

Empirical controlled

- **control group**:
  1% of all global Twitter users → reserve chronological order
- **treatment group**:
  4% of users → personalised recommendations

Personalised recommendations amplify political messages

- right-wing parties benefit more



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

IMPACT OF NAVIGATION SERVICES

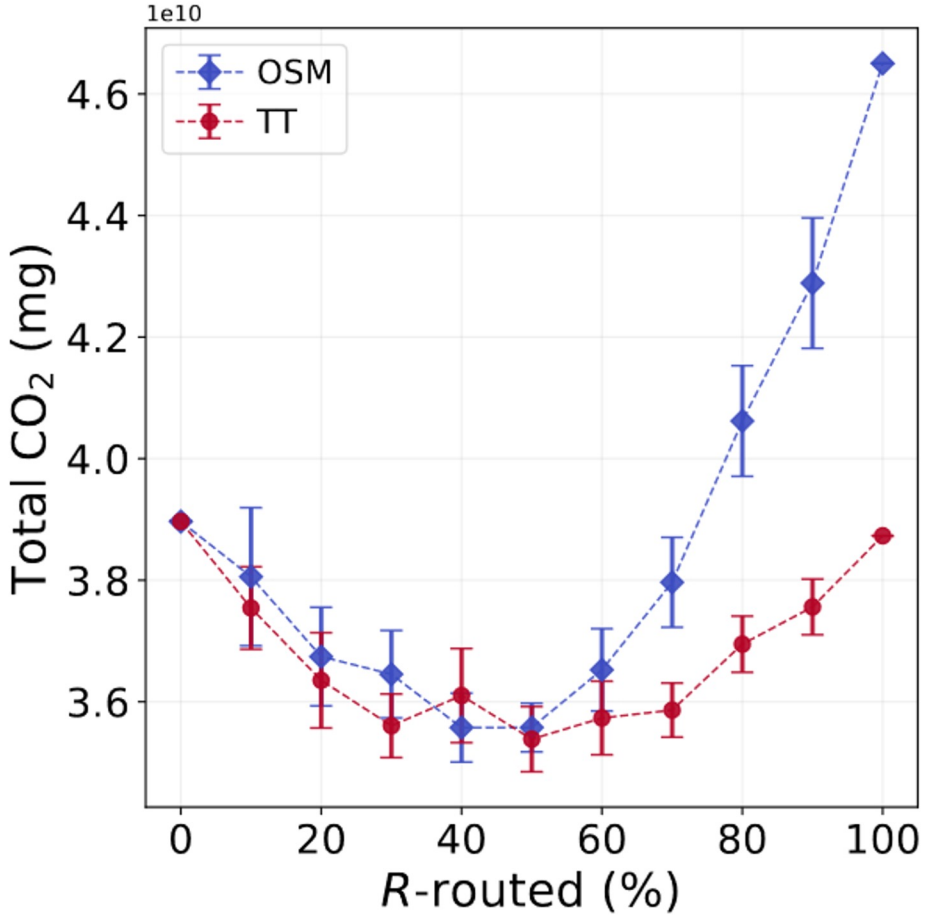
Simulation controlled

Adoption Rate:

- the worst CO2 impact occurs when either all or no vehicles use navigation services

Optimal Adoption:

- reduction in CO2 emissions when approx. half of the vehicles follows routing suggestions



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

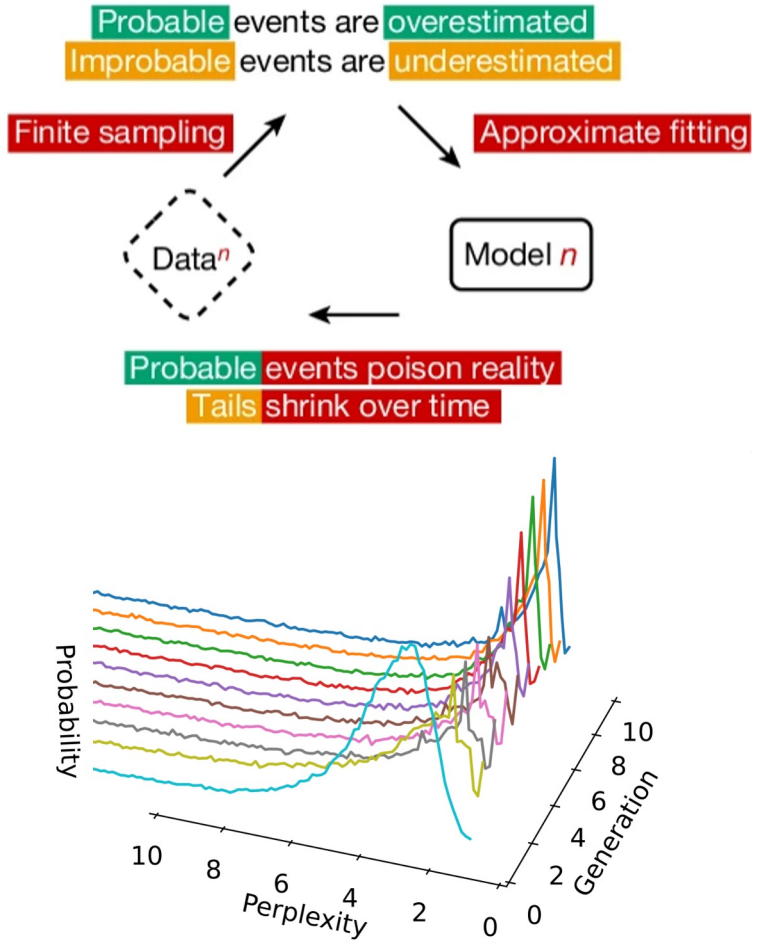
THE CURSE OF RECURSION

Simulation controlled

What happens when LLMs are recursively trained on the synthetic data (**self-consuming loop**)?

Model Collapse:

Decrease in model performance or diversity loss over generations



Shumailov, Ilia, et al. "AI models collapse when trained on recursively generated data." Nature 631.8022 (2024): 755-759.

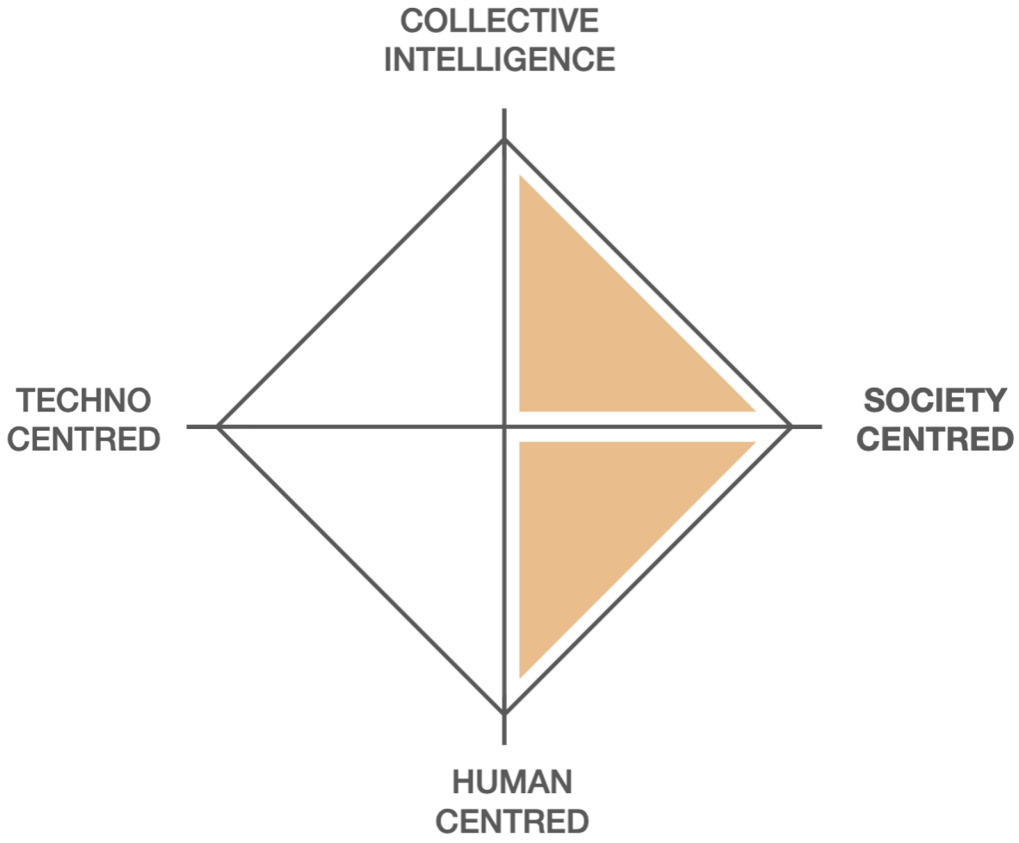
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

HUMAN-AI COEVOLUTION

Field of study, at the intersection between AI and complexity science, which focuses on the *theoretical*, *empirical*, and *mathematical* investigation of the human-AI **feedback loop**

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

A SOCIETY CENTRIC APPROACH



- The feedback loop impacts human well-being also at the societal level
- Controlling the feedback loop requires a new methodological and epistemological approach
- The issues related to human-AI coevolution cannot be solved without legal and political interventions

OPEN CHALLENGES

- Technical challenges
- Epistemological challenges
- Legal challenges
- Political challenges

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

TECHNICAL CHALLENGES

- Methods to **continuously measure the impact** of the feedback loop on the behaviour of humans and recommenders
    - How many iterations might be required before human behaviour substantially changes?
    - How long does it take a generative AI model to collapse?
- Mathematical models to **capture the mechanisms** underlying the feedback loop and its influence on human-AI ecosystems

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

EPISTEMOLOGICAL CHALLENGES

- Understanding the **causal interplay** between humans and recommenders through controlled studies

- Explore **causal relationships by-directionally**: humans and recommenders exert continuous influence on each other, necessitating a holistic study of their co-evolutionary dynamics.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

LEGAL CHALLENGES

- **Limited reproducibility** of studies:
  - Limited access to data for researchers that are external to the platforms managing the recommenders
  - Lack of transparency on how the recommenders are built

- Effective **implementation of legal initiatives** like DSA:
  - it is unclear how vetted researchers will be allowed to access online platforms (Delegated Regulation under definition)

- **Specialized APIs** that allow interacting with platforms
  - to conduct empirical controlled experiments

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

GPAI | THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

Responsible AI for Social Media Governance

A proposed collaborative method for studying the effects of social media recommender systems on users.

Nov 2021 (I edition) Nov 2022 (II edition)

A. Knott, D. Pedreschi et al. Transparency Mechanisms for Social Media Recommender Algorithms: From Proposals to Action. Report, November 2022, Global Partnership on AI

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

INTERNAL STUDIES OF IMPACT

- Social media companies constantly try out different versions of their recommenders on users (A/B test)
- They pick the best ones (by their criteria)
- They use many criteria, but centrally they look for maximization of user engagement

Allen, J and Lawson, A. (2024). On risk assessment and mitigation for algorithmic systems. Integrity Institute report.
https://drive.google.com/file/d/1ZMtZigUcKUq00vakCnbxBCcaA7vaiAix/view

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

THE NEED OF A/B TESTS

- Schemes to run A/B tests would need to be tightly governed
  - Any proposed test would have to be well motivated
  - Proposals would undergo an ethics review
  - Generally, tests would require consent from participating users
- If accepted, A/B tests would need to be publicly pre-registered, to guard against cherry-picking

GPAI international experts believe external researchers should be given the power to run A/B tests under the DSA

MEDICAL ANALOGUE

- Technology in medicine is vetted by rigorous controlled trials, mandated by law

- The results of these trials are openly published:
  - the result is a public science

Why should not we do the same for digital technology with broad social adoption?

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

SOCIO-POLITICAL CHALLENGES

- **Concentration** of "*the means of recommendations*"
  - big-tech companies enjoy a situation of oligopoly
  - recommenders are calibrated to generate profits for the few

- **Lack of political intervention** to redistribute the means of recommendation across a market of many players
  - a more equitable configuration could help develop transparent rules in data access and management of the means of recommendation

> *Change, constant change, inevitable change* is the dominant factor in society today. You can no longer make any reasonable decision without taking into account the world *as it will be*, and this means that you must have a precise intuition of what the world will be like.
>
> Our policymakers, businessmen and ordinary people must assume "**sci-fi thinking**", whether they like it or not, or even whether they know it or not. Only in this way can the terrible problems of today be solved.
>
> Isaac **Asimov**, *My Own View*, The Encyclopedia of Science Fiction, 1978

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

D. Pedreschi, L. Pappalardo, E. Ferragina, et al.
Human-AI Coevolution
arXiv (2024)

C. Wagner et al.

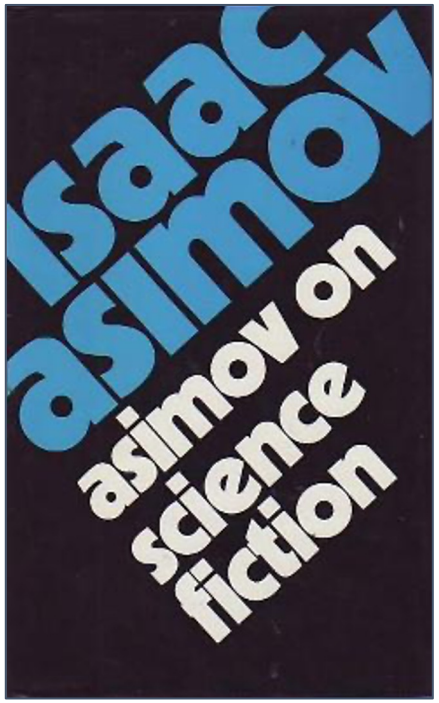
Measuring algorithmically infused societies
Nature 595(7866), 197–204 (2021)

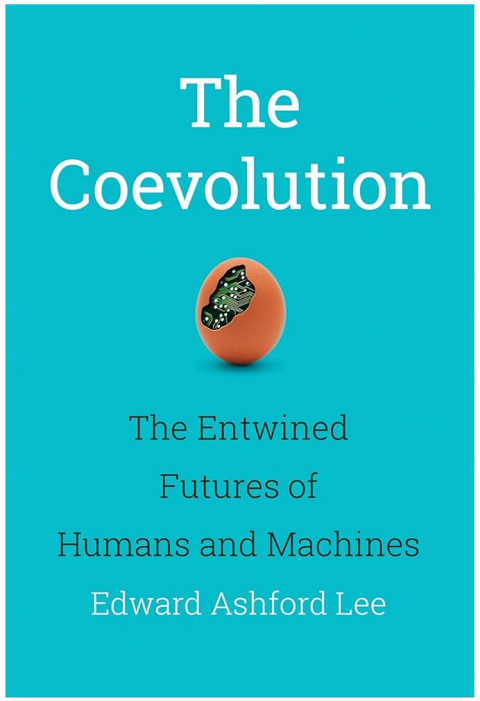
I. Rahwan et al.

Machine behaviour
Nature 568, 477–486 (2019)

Pappalardo, Luca, et al.

A survey on the impact of AI-based recommenders on human behaviours: methodologies, outcomes and future directions.
arXiv (2024).





FUTURE AI RESEARCH

HUMANE AI

SOBIGDATA RESEARCH INFRASTRUCTURE

URBAI