



Project introduction

CHIST-ERA PROJECTS SEMINAR 2021

12-14 April 2021, virtual

Chiara Boldrini

Istituto di Informatica e Telematica
Consiglio Nazionale delle Ricerche
Pisa, Italy

Coordinator: Andrea Passarella, IIT-CNR



Social Explainable AI – overall vision

SAI approach to future ML:

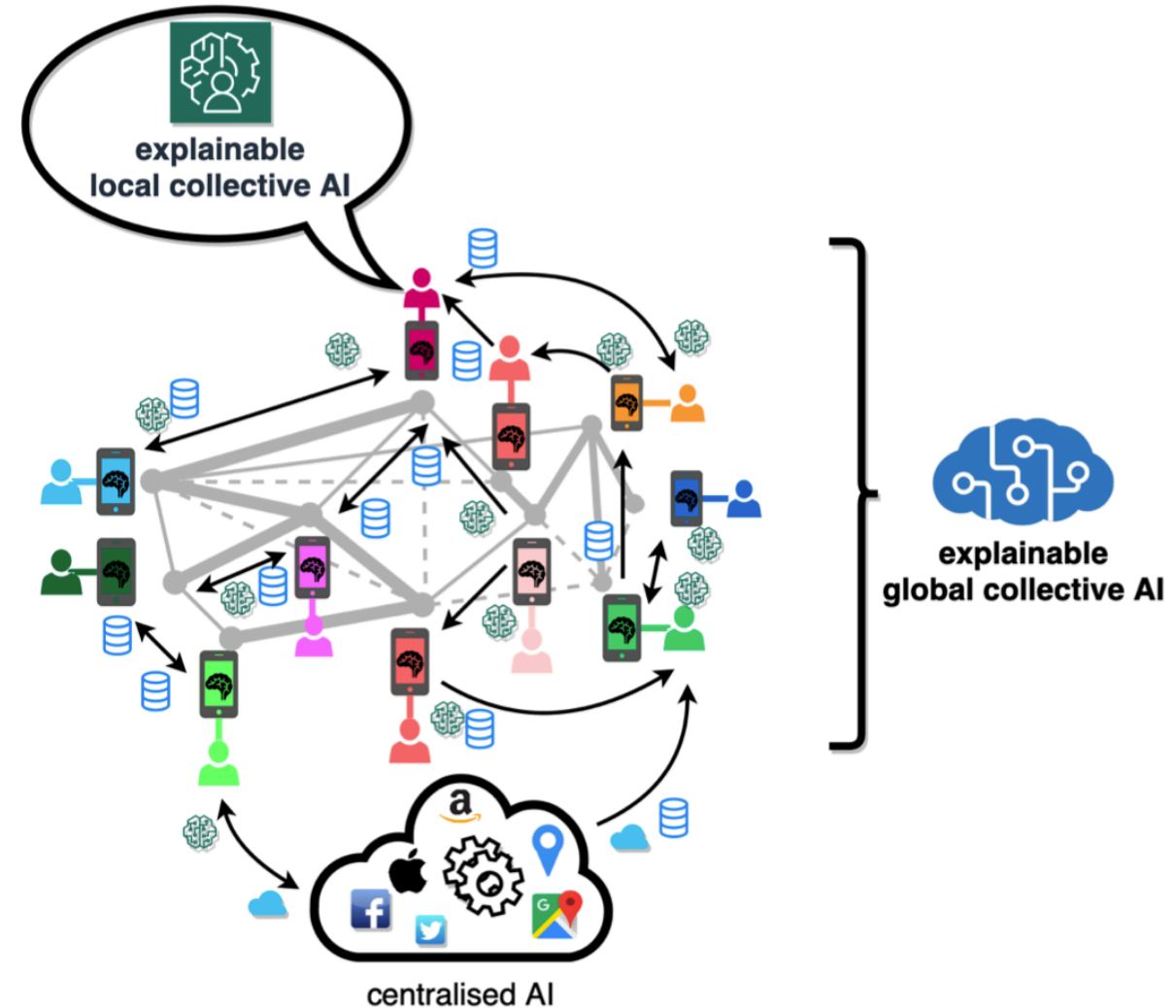
- o From *centralised, black-box* to *decentralised, transparent*

We envision a decentralized collective of personal AI

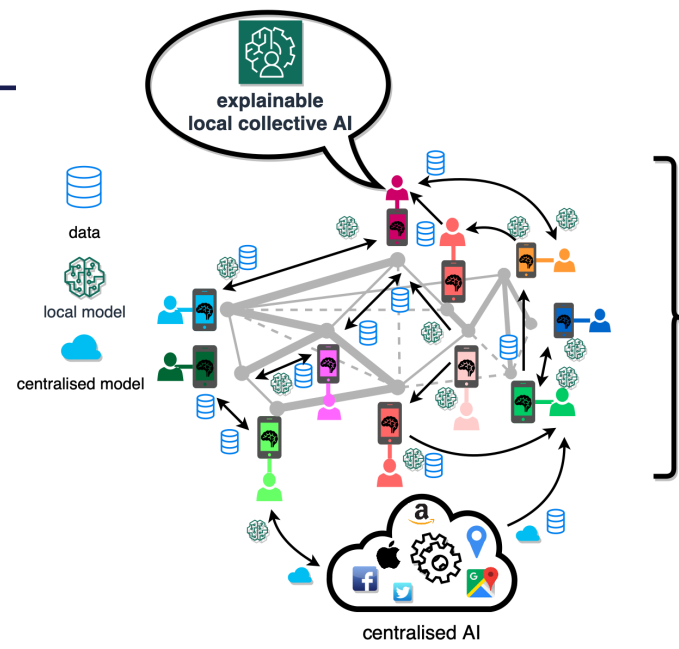
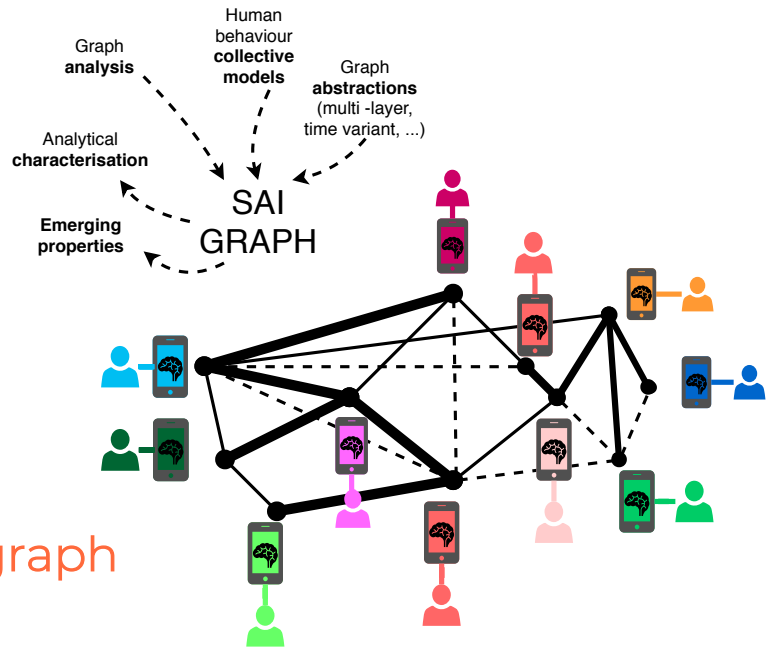
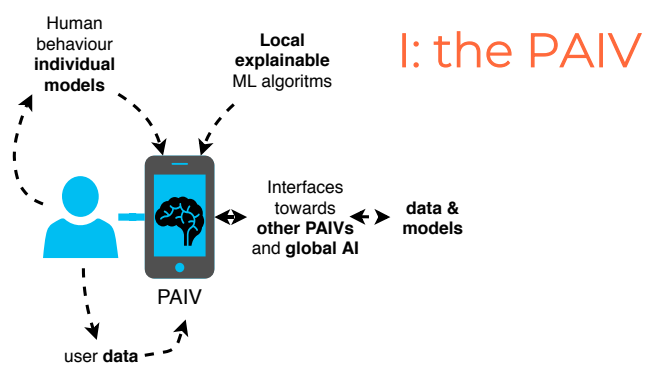
on top of a social network

interpreting data and interacting according to human-centric design principles

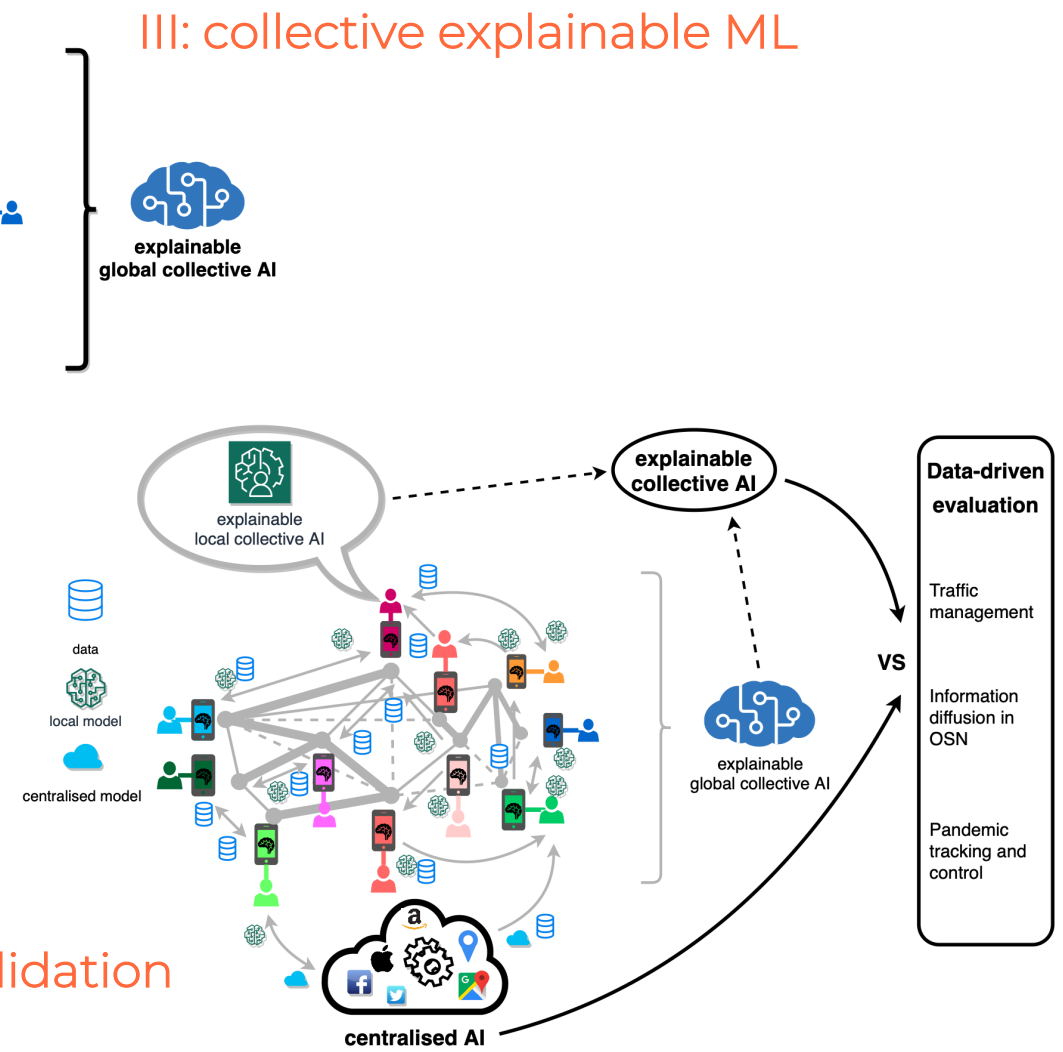
where explainability is guaranteed both at the local and collective level

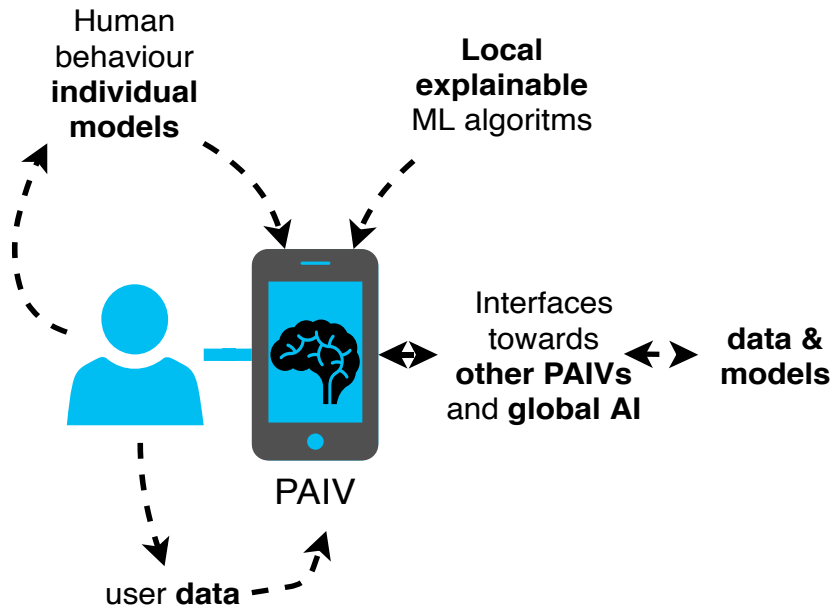


The four pillars

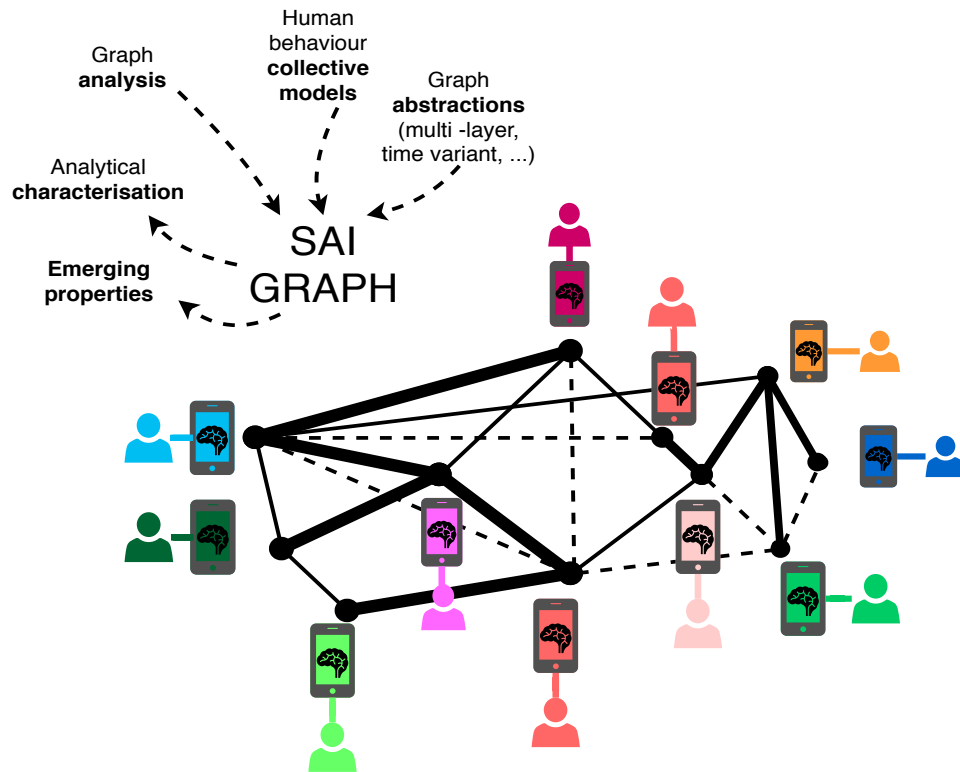


IV: validation





- Human-centric
 - Models of human behaviour
 - Integrated into local ML
- Explainability
 - At the individual level
- From PDS to PAIV
 - Personal data store only deal with data storage/access
 - PAIV also deals with interfacing with other PAIVs
 - Tailored data access



- SAI graph representation

- Multi-layer

- Various facets of human interactions
 - Various communication means
 - ...

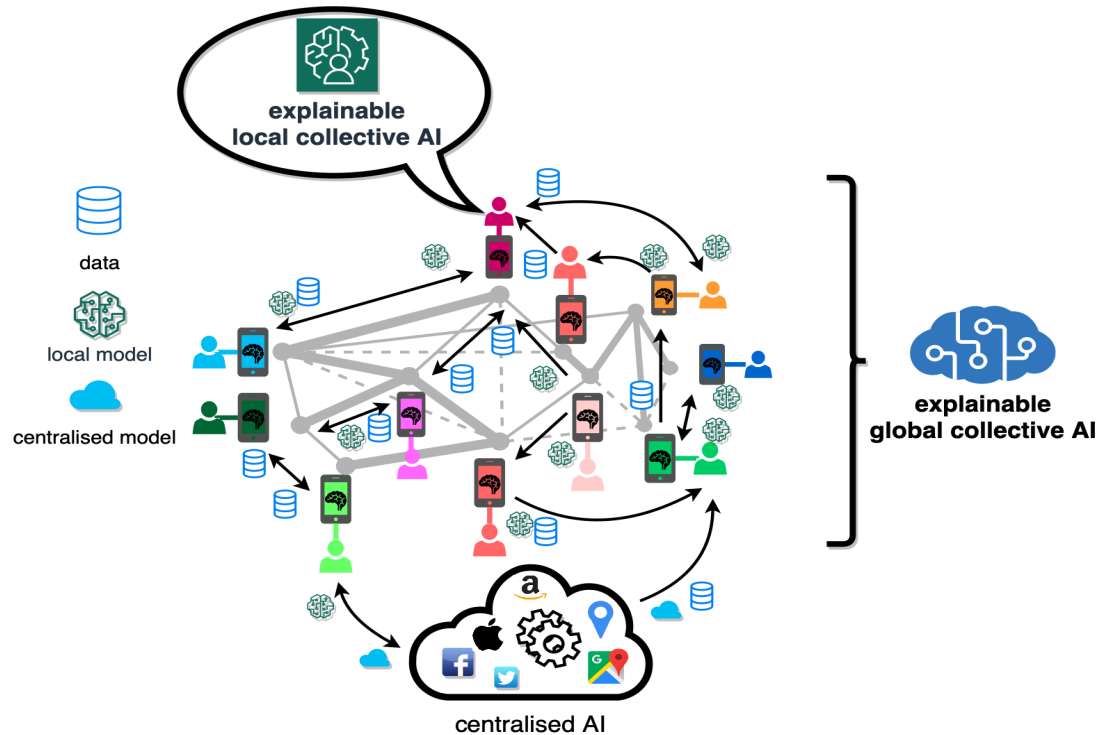
- Dynamic over time and space

- "ephemeral" and contextualised relationships

- Emerging properties

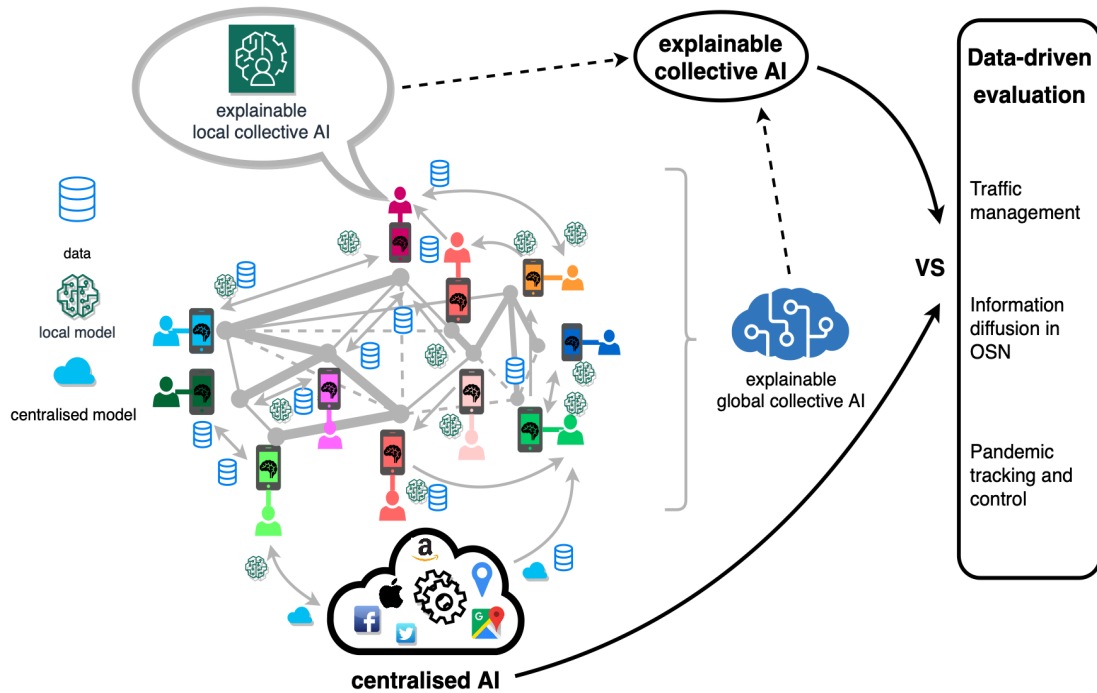
- Macro-, meso-, and micro-scale control points

- "Explainability by graph analysis"



- Human-centric decentralised ML
 - Decentralised aggregation of data/models
- Collective explainability
 - Hierarchies of explanations
- Trust, attacks and robustness
 - “machine learn trust” from users behaviour
 - Impersonation
 - Hijacking of ML models

Fourth pillar – data-driven validation

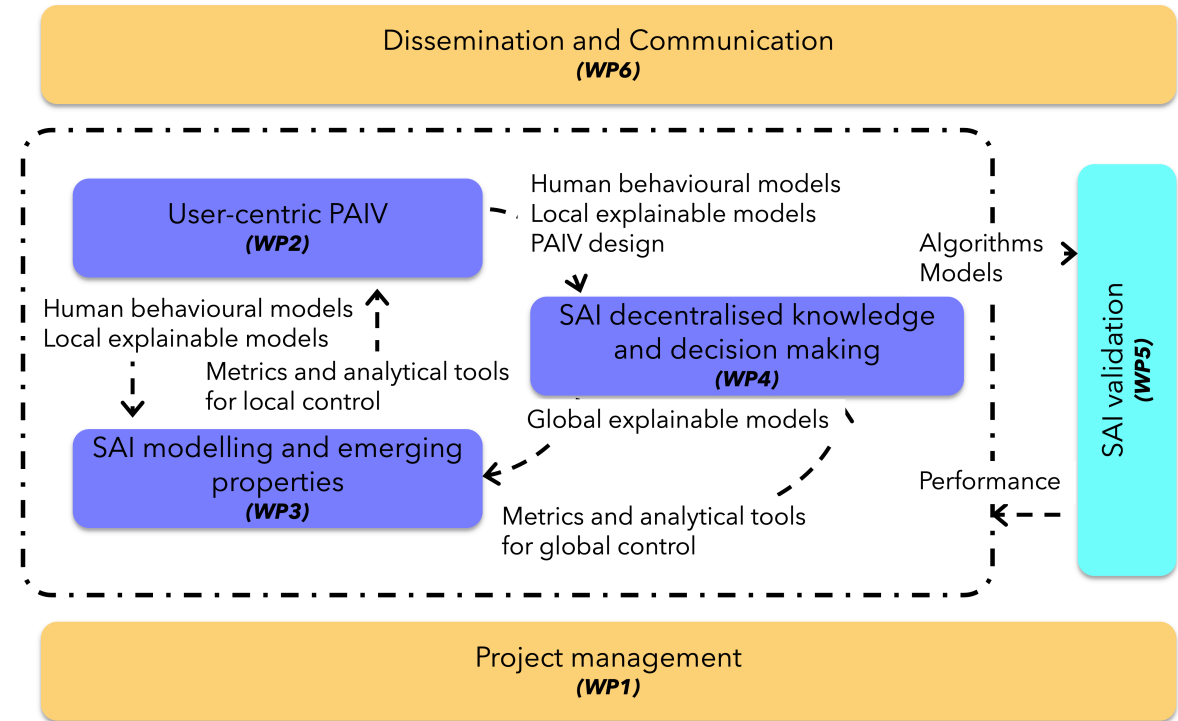


- General method: trace-driven analysis
 - Traces as “centralised as-is”
 - Simulation with SAI as “SAI would-be”
- UC #1: traffic optimisation
 - Personal mobility in/across cities
 - As is: centralised optimiser
 - To be: local decisions based on collective knowledge
- UC #2: online dis-information
 - OSN data
 - As is: info cascades due to centralised optimisation
 - To be: SAI PAIVs decide what to forward and what no; SAI control based on graph properties
- UC #3: pandemic control
 - Synthetic data
 - As is: control via centralised (partial) data sharing
 - To be: control via local (full) data usage

- **Consiglio Nazionale delle Ricerche, Italy**
 - Coordinator
 - Topics:
 - human-centric models
 - Explainable AI
 - Decentralised AI
 - Social networking
- **University of Pisa, Italy**
 - Explainable AI
 - Mobility analysis
 - Complex network modelling
- **Central European University, Austria**
 - Complex network modelling
 - Emerging properties
- **University of Sheffield, UK**
 - BigData analytics and AI
 - NLP
- **Polish Academy of Sciences, Poland**
 - Explainable AI
 - Trust & AI
- **University of Tartu, Estonia**
 - Complex network modelling
 - Social networking
- **Sofia University “St. Kl. Ohridski”, Bulgaria**
 - ML models
 - Explainable AI

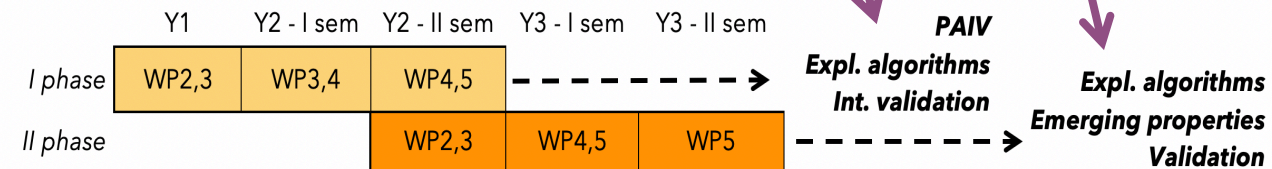


- PAIV (WP2)
 - Human models
 - Local explainable ML algorithms
 - PAIV interfaces for sharing
- SAI Graph (WP3)
 - Control points for decentralised optimisation
- Decentralised ML (WP4)
 - Decentralised algorithms
 - Collective explanations



- Phase I (~M1-M18)
 - Human behavioural models
 - PAIV & local explainable ML
 - SAI graph characterisation
 - Initial decentralised ML
 - Initial evaluation
- Phase II (~M18-M36)
 - SAI emerging properties & control
 - Decentralised ML
 - Collective explanation
 - Validation

	Y1				Y2				Y3			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WP1: Project Management T1.1: Administrative and technical management	D1.1	D1.2					D1.3				D1.4	
WP2: User-centric PAIV T2.1: Individual and social behavioural models T2.2: Explainable local AI models T2.3: From PDS to PAIV	D2.1	D2.2							D2.3			
WP3: SAI modelling and emerging properties T3.1: The SAI graph T3.2: SAI emerging properties			D3.1	D3.2					D3.3			
WP4: SAI decentralised knowledge and decision making T4.1: Decentralised human-centric ML models and decision-making algorithms T4.2: Collective Explainable AI models T4.3: Trust and malicious behaviours							D4.1	D4.2				
WP5: SAI Validation T5.1: Tools for data collection, curation, storage and sharing T5.2: Validation use case #1: traffic optimisation T5.3: Validation use case #2: (dis-)information diffusion in OSN T5.4: Validation use case #3: pandemic tracking and control							D5.1/2				D5.3	
WP6: Dissemination and Communication T6.1: Dissemination via publications and liaising with related projects T6.2: Dissemination via events and communication T6.3: Training	D6.1	D6.2					D6.3				D6.4	
Milestones		M1	M2		M3		M4			M5	M6	



- Scientific
 - lay the foundations for future human-centric decentralised AI systems
- Technological
 - supports the booming expansion of AI, adding the dimensions of decentralization and human centrality
- Societal
 - the awareness and deep control over personal data that SAI provides to the users can have beneficial effects that go beyond the individual
 - e.g., enabling the collaborative objectives that are today obstructed by the lack of proper protocols, trust and guarantees on data usage



Thank you for your attention!

“The ultimate goal of SAI is to provide the *foundational elements* enabling a *decentralised collective of explainable PAIVs* to evolve *local and global AI models*, whose processes and decisions are *transparent, explainable* and *tailored* to the needs and constraints of *individual users*”

