Allegato B – Descrizione del progetto

Introduction to the FAIR Project

The Coordinated Plan on Artificial Intelligence (launched in 2018 and revised in April 2021[1]) puts forward a concrete set of joint actions for the European Commission and Member States on strengthening Europe's leading position in the development of human-centric, sustainable, secure, inclusive and trustworthy AI. The European Commission also released a legal framework on AI[2], which addresses the risks of AI and proposes proportionate and flexible rules to address the specific risks posed by AI systems, setting the highest standard worldwide. The National Strategic Program on AI (2022-2024)[3], released by the Italian government in November 2021 coherently with the EU strategy on AI and the Intervention Area 4 of the Italian PNR (2021- 2027)[4], sets for Italy the ambitious goal of becoming a global research and innovation hub of AI. The plan is to: i) advance frontier research in AI, both for fundamental and applied research, ii) reduce AI research fragmentation, iii) develop and adopt human-centered and trustworthy AI in the public and private sector, iv) increase AI-based innovation and development of AI technology, v) develop AI-driven policy and services in the public sector, and vi) create, retain and attract AI talent.

Over the course of the last decade, AI researchers have made groundbreaking progress in hard and longstanding problems related to machine learning, computer vision, speech recognition, and autonomous systems. Despite the success of AI, its adoption so far is mostly in low-risk applications, while the uptake in medium/high-risk applications, which might have a

This proposal for an extended partnership titled Future Artificial Intelligence Research (hereafter FAIR) is the response of the Italian AI scientific community to the National Strategic Program. FAIR takes on the challenge to set the agenda of frontier research for the AI methodologies and techniques of tomorrow.

deeper transformative impact on our society, such as in healthcare, public administration, safety-critical industry etc., is still low compared to expectations. The reasons for such lagging are profound. Adoption barriers include perceived challenges to the autonomy and the oversight capacity of human users, the required effort, dissatisfaction with user interfaces and, above all, trust concerns related to poor users' knowledge about the assumptions, limitations and capabilities of AI systems. Lack of trust undermines the spread of innovations.

FAIR AMBITION: Well beyond currently available technologies, we need AI systems capable of interacting and collaborating with humans, of perceiving and acting within evolving contexts, of being aware of their own limitations and able to adapt to new situations, and interact appropriately in complex social settings, of being aware of their perimeters of security and trust, and of being attentive to the environmental and social impact that their implementation and execution may entail. In short, **we need an AI that does not yet exist.**

[1] European Commission, Coordinated Plan on Artificial intelligence 2021 Review, COM(2021)

[2] European Commission, Proposal for a Regulation laying down harmonized rules on artificial intelligence, (COM(2021) 206 final

[3] https://www.mur.gov.it/sites/default/files/2021-12/AI_strategia_nazionale_programma-strategico-eng.pdf

[4] https://www.mur.gov.it/it/aree-tematiche/ricerca/programmazione/programma-nazionale-la-ricerca

This is why fundamental, multidisciplinary research is necessary to shape the AI of the future. FAIR envisions new forms of AI that are:

1. *Human-centered*: co-evolve with the human "in-the-loop", at individual and collective level.
2. *Integrative*: bridge across the different AI methods, technologies, disciplines, and competences.
3. *Resilient*: operate in challenging, noisy, uncertain real-world settings.
4. *Adaptive*: perceive, learn and act in dynamically changing environments.
5. *High-quality*: meet quality standards for high-risk, high-accuracy, safety-critical applications.
6. *Symbiotic*: promote effective human-machine interaction and collaboration.
7. *Edge/exascale*: operate at the infinitely small on the edge, and at the infinitely big on the cloud.
8. *Pervasive*: operate ubiquitously in different social settings.
9. *Green-aware*: consider the environmental dimension "by-design".
10. *Sustainable and bio-cognitive*: mimic the principles of biological systems at multiple scales.

While these 10 themes will be the topics of 10 large spoke-projects, a number of fundamental scientific and technological challenges will be tackled with coordinated interspoke actions, called *transversal projects - TP*, where each involved spoke will contribute from its own specific perspective:
(TP1) *Legal and Ethical Design of Trustworthy AI Systems*: how to create responsible, trustworthy AI "by- design", "in-design" and "for designers".
(TP2) *Vision, Language and Multimodal Challenges*: how to create AI agents capable of perception in real, complex environments with multiple combined modalities (text, speech, images, video, …)
(TP3) *Learning and Reasoning from Individual to Communities to Society*: how to create AI agents that integrate learning and reasoning to assist decision making at multiple scales (individual, societal)
(TP4) *Adjustable Autonomy and Physical Embodied Intelligence*: how to create autonomous AI systems capable to understand the limits of their autonomy, asking for human supervision when appropriate.
(TP5, TP6) *Frontiers of Machine Learning*: how to gear the methods of mathematics and physics to understand why and when machine learning works, and how to expand the frontiers of "lifelong", continual, incremental learning and meta learning (learning to learn).
(TP7) *Data Centric AI and Infrastructures*: how to manage, prepare and curate large, high- quality data for AI development.

The figure highlights the synergies among Spokes and Transversal Projects. The paradigm of TPs is a dynamic concept aimed at amplifying the synergies among spokes, providing the "*humus*" for further collaborations, so that, in the course of the program, new additional cross- cutting challenges might be envisaged and new TPs will possibly emerge.

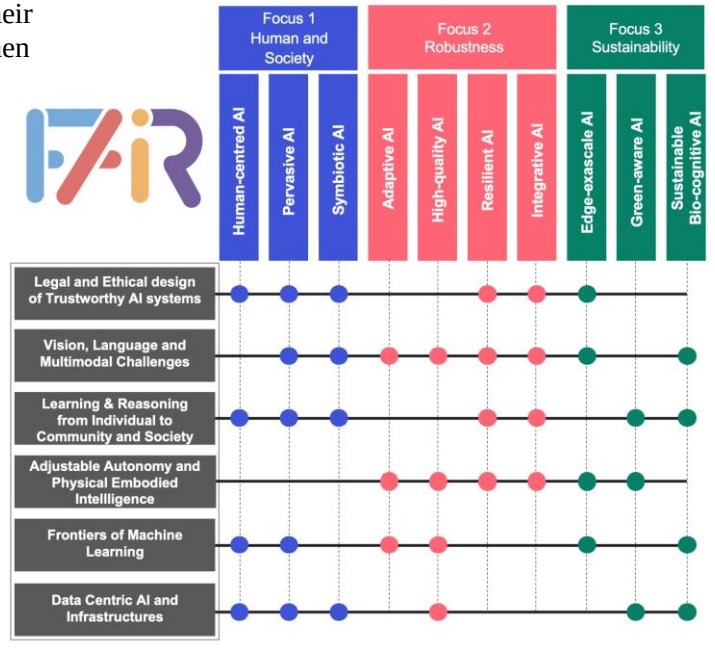

Fig A.1 Spokes and Transversal projects

While pursuing its ambitious scientific goals, FAIR intends to contribute significantly to the objectives set forth in the Italian Strategic Programme on AI. In short:

FAIR is organized according to the 'Hub & Spokes' model, which is extensively presented in Section B. The coordinator is the National Research Council (CNR) and the programme is implemented by the functional merge of activities planned by the spokes in charge of specific thematic areas. The spokes have all the same structure: i) 5-8 research WPs that are full-fledged projects. The WPs execution can be partially supported by cascading calls; ii) a WPi.0 responsible for the overall management and coordination, and with the special task of contributing to a Research and Innovation (R&I) roadmap for the present and future AI and to the dissemination plan in charge of the Hub. The last WP of any spoke concentrates on experiments/case studies/pilots in challenging domains. Moreover, each research WP is also responsible for boosting the curation and publication of its results (data, methods, systems, use cases, workflows) into the FAIR project catalog of the FAIR platform supported by the Hub.

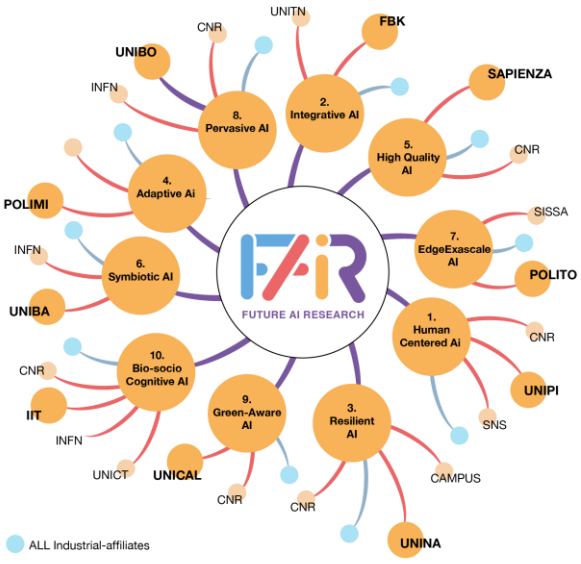

Fig. A.2 Spokes and Affiliates

The FAIR model of cooperation is based on Transversal Projects (TPs), which sustain strong interspoke interactions among the participating research groups. TPs bring together research on the same challenge that spans across the spokes. For each spoke, broad areas of research activity and the relative methodology are hereafter briefly summarized.

Spoke 1: HUMAN-CENTERED AI. PI: Dino Pedreschi, co-PI: Francesco Marcelloni - CNR, UNIPI, SNS. Spoke 1 focuses on the following three research activities. (i) **Human-in-the-loop** machine learning and reasoning aim at allowing humans to understand and steer the learning and reasoning of AI systems, interact synergistically and co-evolve with ML models. Specific goals include explainable AI, neuro-symbolic learning and lifelong learning. (ii) **Social-aware AI** aims at understanding and governing the societal outcomes of large- scale, networked socio-technical systems of humans and AIs, e.g., social media and online marketplaces. Specific goals include modeling such systems and decentralized networked learning. (iii) **Design of trustworthy AI systems** aims at responsible (co-)design, development, validation and use of trustworthy AI systems, including certification, to incorporate "by-design" European laws and ethical values. A strong multidisciplinary focus pervades Spoke 1, with synergies between AI and human-computer interaction, cognitive sciences, complex systems, math, ethics, law, and social sciences.

Spoke 2: INTEGRATIVE AI. PI: Paolo Traverso, co-PI: Paolo Giorgini - FBK, UNITN.
Spoke 2 proposes an overarching interdisciplinary approach to Integrative AI, with integration along two dimensions. *Horizontal* integrative-AI deals with the combination of several high-level elements: (i) **modeling**, **reasoning**, and **deliberation** in autonomous agents; (ii) **interaction** with the environment, including humans; (iii) **cooperation** among multiple artificial and human agents. *Vertical* Integrative AI is oriented to theories and techniques combining different scientific methods, competences, and technologies in specific AI sectors. This includes the integration of heterogeneous computational representation and reasoning/learning techniques, e.g. **data-driven and model-based approaches, symbolic and sub-symbolic representations**, as well as the integration of research closely related to AI such as formal methods, complex systems, software engineering, and cognitive science. `

Spoke 3: RESILIENT AI. PI: Carlo Sansone, co-PI: Giuseppe Di Pietro - UNINA, Campus, CNR.

Spoke 3 addresses the study of AI foundational methodologies that are aimed at processing data in-the-wild, making the performance of AI resilient and robust in challenging contexts. Spoke 3 studies how learning algorithms can cope with the problem of **training with real-world data** and we devise novel theories, methods, and automated instruments to address the current limitations of AI-intensive software system development, and also pay attention to the ethical and legal issues that involve AI applications in-the-wild. The research activities to be carried out include: (i) the definition of appropriate **data augmentation techniques**, when data are incomplete or not adequately representative, while analyzing, monitoring, and improving the fairness of the machine learning algorithms; (ii) the definitions of algorithms that are both **resilient and robust with respect to possible external attacks** (also deriving from training with "malicious" data); (iii) the investigation of the implications related to the design, validation & verification, evolution and operation of the **software that implements machine or deep learning algorithms**, when they have to work in-the-wild; (iv) the **ethical and legal issues** connected with the use of real-world data.

Spoke 4: ADAPTIVE AI. PI: Nicola Gatti, co-PI: Nicolò Cesa-Bianchi - POLIMI, Bocconi.
Spoke 4 aims at a **unified theory combining observational and interventional learning,** that provides formal models for adaptivity. Natural questions to answer in this challenge concern when some specific property can be guaranteed and under which conditions. The scientific question we aim at answering is how can one adapt best to a given change and which guarantees one can provide in a specific setting. To obtain rigorous theoretical guarantees on the behavior of adaptive systems, this area of research combines algorithmic design with mathematical tools from many disciplines, including (i) optimization, control theory, and dynamical systems theory; (ii) the complex systems viewpoint of statistical physics and

general computational paradigms, such as quantum computing; (iii) the connection to computational neuroscience.

Spoke 5: HIGH-QUALITY AI. PI: Giuseppe De Giacomo, co-PI: Luca Iocchi - Sapienza, CNR.
Spoke 5 aims at characterizing the various **qualities of interest in AI-based systems**, their **measurability**, **assessment** and **certifiability**, across the whole AI. Qualities of interest are broad, and include: **foundational and mathematical qualities**, **engineering qualities**, **social and ethical qualities**, **legal and certifiable qualities**, **qualities that concern sustainability and economic value,** and **scientific qualities**. We aim to develop methodologies for designing and engineering AI systems focused on such qualities, their measures and assessments. We consider AI systems that are able to self-assess the qualities of their outcomes, and turn to human supervision if they do not meet the required standards. The kind of **services** we want to guarantee wrt qualities are: assessment, certification, measurement, and evaluation. Finally, we address AI systems that are able to self-assess the qualities of certain decisions, conclusions that they draw, and be ready to turn to human supervision if the qualities of interests do not meet suitable standards.

Spoke 6: SYMBIOTIC AI. PI: Donato Malerba, co-PI: Giovanni Semeraro - UNIBA, INFN.
Fostering human-AI symbiosis requires designing AI systems by leveraging the techniques developed within the **Human-Computer Interaction** (**HCI**) community, taking into account user needs and desires, as well as emotions, intentions, actions, etc. This spoke lays the foundations of Symbiotic AI (SAI) along with three main goals: 1) enabling **SAI system design** with principles and methodologies of HCI; 2) enabling **SAI development** through both data-driven and model-driven approaches to endow AI systems with human understanding capabilities and improve AI systems' performance with user input; 3) enabling **SAI applicability** by improving the understandability of AI systems, enhancing the acceptability of SAI systems, supporting the sustainability of SAI, and providing the necessary infrastructure.

Spoke 7: **EDGE and EXASCALE AI.** PI: Barbara Caputo, co-PI: Tatiana Tommasi - POLITO, SISSA.
Spoke 7 aims at designing and studying **new AI algorithms** that can leverage by construction over the intrinsic properties of edge and exascale hardware, articulated over three main challenges. The first is how to adapt existing artificial intelligence algorithms to the specific **requirements of the edge and exascale computational frameworks**. The second challenge is how to design a new generation of AI algorithms able to retain all the characteristics of existing state-of-the-art methods, while being **optimized by design for these two specific hardware**. The third challenge is how to design **innovative edge and exascale AI algorithms that are sustainable**.

Spoke 8: PERVASIVE AI. PI: Michela Milano, co-PI: Rita Cucchiara - UNIBO, CNR, INFN.
To make AI positively pervasive, it is necessary to holistically address challenges related to the **robustness and accuracy** of models, the **efficiency of implementations**, and the **cultural acceptability and sustainability** of AI. A final aspect that is crucial in this pervasive perspective is the **education aspect** that, on one hand, should increase awareness and engagement toward AI systems, lowering barriers and improving uptake, and on the other hand should improve educational processes with AI tools and solutions. Spoke 8 addresses the **challenges of pervasive AI** by considering ubiquitous intelligent objects, socio-technical systems, perceptive and multimodal systems, by seamlessly gluing interacting model components: artificial infrastructures and intelligent objects, social dynamics and impact, human behavior, environmental impacts, legal compliance, economic aspects and acceptance/awareness/trust in intelligent systems.

Spoke 9: GREEN-AWARE AI. PI: Nicola Leone, co-PI: Francesco Scarcello - UNICal, CNR.
Spoke 9 conducts foundational research to define a novel generation of AI systems and algorithms, in

which - no matter the specific application domain being considered – **the green dimension is considered as a first-class citizen** in order to mitigate the risks defined above. In particular, Spoke 9 focuses on the following methodological challenges: **KR&R frameworks for green-aware AI, interactions among green-aware agents, green AI, green-aware XAI, adjustable green-aware AI.** In addition, Spoke 9 targets AI applications that address **AI for Green**: environmental challenges that have a broad impact on the life of the citizens, with domains such as renewable energies, circular economy, environmental monitoring and risk assessment, tourism and agritech.

Spoke 10: SUSTAINABLE BIO-SOCIO-COGNITIVE AI. PI: Massimiliano Pontil, co-PI: Concetto Spampinato. IIT, UNICT, CNR, INFN.
Making **sustainable bio-socio-cognitive** virtual and physical agents requires rethinking current AI systems, and rests on highly interdisciplinary foundational research domains: (i) **cognitive science and neuroscience- informed machine learning**; (ii) **efficient, robust and informed machine learning, (iii) computational reasoning models and experimental/social psychology** identifying human factors and social dynamics for an effective integration of AI within human societies. Spoke 10 aims at developing the following research activities: neuro-bio-inspired learning paradigms encoding high-level and low-level mechanisms of the human brain, multisensory and cross-modal learning, egocentric perception to see and act, Efficient and informed machine learning, optimization and robustness in machine learning, Exploring the sense-perceive-plan-act loop in embodied virtual and physical robots, grounded world models exploiting prior knowledge and experience, hybrid collective intelligence.

Description of SPOKE 2: Integrative AI

Objectives: The objective of Integrative-AI is to overcome the barrier of complexity that prevents addressing real-world situations involving autonomous situated agents, their interaction with the environment, and the cooperation among them and with humans. We will investigate the integration of heterogeneous computational representation and reasoning techniques, such as the integration of data driven and model based approaches, symbolic and sub-symbolic representations. We will develop unifying theories and frameworks, novel tools and techniques, and provide theoretical and experimental evidence of the advantages of Integrative-AI to address three main challenges of AI: (i) modeling, reasoning, and deliberation (WP2.1, 2.2, and 2.3); (ii) interaction with the environment, including humans (WP2.4, 2.5, 2.6, 2.7); (iii) cooperation among multiple artificial and human agents (WP2.8). We call this first grade level of Integrative-AI, the vertical level of Integrative-AI. Moreover, most research in AI works on these three main challenges separately. However, most often one is dependent on the other. We call this second level of Integrative AI, the horizontal level of Integrative-AI (WP2.9).

Description of Spoke 2 Work PACKAGES

- **WP2.0 – Management, cascading calls, strategic research and innovation roadmap**
- **WP2.1 Integrative-AI for Verification, Synthesis, and Autonomy -** We will research algorithms and tools for automated reasoning, optimization, verification and validation, synthesis and planning for highly expressive logical formalisms. The purpose is to increase automation and efficiency by integrating heterogeneous symbolic and subsymbolic approaches to AI. The foundations of trustworthy autonomous architectures will be investigated. *Connected to TP Adjustable Autonomy and Physical Embodied Intelligence (TP4)*

- **WP2.2 – Models for integrative AI** - Development of well-founded approaches for the integration of heterogeneous AI models (including neural, probabilistic, logical, dynamic and decision models) into a single agent; study of the dynamic of these models due to the interaction of the agent with the environment and the intra-agent integration that happens at the social level.
- **WP2.3 – Multi-Perspective Knowledge Representation** - We will research formalisms and algorithms to integrate the multiple perspectives of a real-world domain into holistic formal representations, and to discover, align, co-evolve, and exploit implicit and explicit knowledge. This will empower AI systems with the ability to build, make explicit, use, and share formal representations of the world.
- **WP2.4 - Integrative-AI for Multi-Modal Perception** - We will investigate new paradigms and architectures for cross-modal content generation, understanding, and searching by considering continual and lifelong learning strategies and dealing with scarce labeled data and multimodal fusion strategies and by studying and implementing novel approaches for deep network distillation, compression, pruning and quantization. *Connected to TP Vision, Languages in Multimodal Challenges (TP2).*
- **WP2.5 –Integrative AI for Natural Language Understanding** - We will tackle fundamental challenges related to Integrative AI for Natural Language Understanding. The state-of-the-art neural models will be able to integrate knowledge from heterogeneous sources, will be well integrated with societal needs, adapted into realistic and rich communicative contexts, and combined to obtain comprehensive models. C*onnected to TP Vision, Languages in Multimodal Challenges (TP2).*
- **WP2.6 – Integrative-AI for human-computer interaction** - The WP addresses basic research topics in human interaction by adopting a multidisciplinary perspective which includes machine learning, computational linguistics, cognitive & social psychology. The fundamental goal is to identify and design the mental models for AI systems to be accepted and made adaptable to different user groups.
- **WP2.7 – Integrative AI for Embodied Systems** - We will focus on the foundations of an embodied AI i.e., AI systems working in the physical world. We will create systems combining adaptability with safety, flexibility with reliability. Our solutions will rely on a synergistic combination of model-based techniques with learning from data and humans. *Connected to TP Adjustable Autonomy and Physical Embodied Intelligence (TP4).*
- **WP2.8 – Cooperative and hybrid human-machine intelligence** - The WP will develop the theoretical and methodological foundations for hybrid human-machine learning and decision making and collective intelligence in hybrid human-machine networks. It will devise AI models for explainable interactive human-machine learning and decision making, and develop socially-inspired cooperative AI systems and systems combining humans and AI agents. *Connected to TP Learning and Reasoning from individual to communities to society (TP3) and TP Legal and Ethical Design of AI (TP1).*
- **WP2.9 – Horizontal level of Integrative-AI, experiment case studies/use cases/pilotsPlatform extension, Synergies Industry in challenging domains** - Coordination among Work Packages 1-8 by addressing the challenge of the horizontal Integrative-AI. Set up of experimental evaluation through collaboration with private companies and public

institutions in real world settings, such as autonomous systems for industrial plants, applications for social aware systems, and virtual coaching for health and wellbeing.